

BIDIRECTIONAL INFLUENCES OF SOCIAL FEEDBACK  
ON PARENT-INFANT COMMUNICATION

A Dissertation

Presented to the Faculty of the Graduate School  
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy

by

Rachel Ryan Albert

August 2013

© 2013 Rachel Ryan Albert

BIDIRECTIONAL INFLUENCES OF SOCIAL FEEDBACK  
ON PARENT-INFANT COMMUNICATION

Rachel Ryan Albert, Ph.D.

Cornell University 2013

How does parent-infant interaction impact infant's communicative development? The aim of the studies presented here is to better understand the function of early social interactions in providing opportunities for vocal development and language learning. Parental behaviors that organize infant attention can create socially coordinated, *aligned* interactions in which attentional focus, object movements, and labeling are synchronized in ways that promote word learning. Alignment is a process in which communicators' production and comprehension of speech become coupled to establish common ground. By examining language development as an outcome of a socially embedded system of bidirectional feedback, the experiments presented in this thesis investigates how prelinguistic infants and their caregivers develop vocal alignment to support the emergence of mature communication. The research presented here: (1) investigates the constraints on socially guided statistical learning in influencing infants' learning of new phonological patterns from caregivers contingent responses to infant vocalizations; (2) determines the function of the infant's own object-directed prelinguistic vocalizations in facilitating the mapping of word-object associations; (3) examines the role of caregiver experience in responding to infant vocalizations; and (4) assesses the characteristics of infant vocalizations that elicit social interactions from caregivers. Taken together, the results of these experiments enhance our understanding of the cognitive mechanisms by which communicative alignment influences infant language development.

## **BIOGRAPHICAL SKETCH**

Rachel grew up in Northern New Jersey and graduated from Kinnelon High School in 2004. She went on to undergraduate studies at Ohio Wesleyan University (OWU) where she double majored in Psychology and Sports Science. Her academic interests focused on the study of child development. Rachel's senior honors thesis, supervised by Dr. Kim Dolgin, was titled "Lasting effects of short-term training on preschoolers' street-crossing behavior". In May 2008, Rachel graduated from OWU magna cum laude, Phi Beta Kappa, and with university and Psychology honors.

Rachel came to Cornell University in 2008 to begin graduate studies working with Drs. Michael Goldstein and Jennifer Schwade in the Behavioral Analysis of Beginning Years Laboratory (BABY Lab). While in the BABY lab, Rachel supervised many undergraduates and mentored several students on independent research projects and honors theses. During her time at Cornell, Rachel also sought out several independent teaching experiences including teaching Writing in the Majors, seminars on Infant Development, and Introduction to Psychology at the Auburn Correctional Facility as part of the Cornell Prison Education Program. After the completion of her PhD, Rachel will join the faculty of the Psychology Department at the University of Wisconsin Stevens Point.



This document is dedicated to Daniel and Noah  
for your unconditional love and support

## **ACKNOWLEDGEMENTS**

First, I'd like to thank my academic advisors Drs. Michael Goldstein and Jennifer Schwade for their endless guidance and support the past five years. I would also like to thank my committee members: Drs. Steven Robertson, David Smith, James Cutting, and Shimon Edelman. Your thoughtful questions, advice, and encouragement were instrumental to conducting and completing the research presented in my dissertation.

Second, I'd like to thank each member of the BABY Lab, past and present. Without you, this work could never have been conducted. Particular thanks to the research assistants who went above and beyond in helping recruit, test, and code the participants in the studies presented here: Jessica Mesick, Taylor Udell, Frederique Verly, Brenna McGuire, Emma Fitzpatrick, Stephanie Ellman, Sarah Edmunds. Your hard work and dedication to my projects was appreciated more than you could ever know. I would also like to thank Oti Menyhart for helping to push me through each disappointment and celebrating every accomplishment. Thank you also to the families of Ithaca and Syracuse who made this research possible. It was a joy meeting your babies and learning about their development!

Finally, I'd like to express the deepest appreciation to my family for your unwavering support and encouragement. Thank you Mom and Bruce for the opportunities that brought me here and all of your love along the way. Thank you Hayley for bringing humor to every situation. Thank you to my parents and Carol and Ron for the countless visits to give us a break – especially all of the babysitting this past year. To our greatest joy, Noah: your arrival provided balance, perspective, and motivation that I never could have imagined. Thank you for the smiles and kisses every time I get home. Above all, I'd like to thank Daniel. Your daily love, reassurance, and support made this dissertation possible.

## TABLE OF CONTENTS

Abstract	i
Biographical Sketch	ii
Dedication	iii
Acknowledgements	iv
Table of Contents	v
List of Figures	vi
List of Tables	vii
Chapter 1: Introduction	1
Chapter 2: The ‘aha’ moment: What is learned when infants learn to produce new phonological regularities?	24
Chapter 3: Playback of prelinguistic object-directed vocalizations facilitates word learning	49
Chapter 4: Determinants of caregiver responding to infant vocalizations: Findings from the playback paradigm	79
Chapter 5: Conclusions	136

## LIST OF FIGURES

Figure #	Figure Caption	Page #
2.1	Mean frequency of target VCV disyllables (e.g. V[h]V or V[g]V) and non-target VCV disyllables by period ( $\pm 1$ SE).	36
2.2	Mean proportion of target VCV disyllables (e.g. V[h]V or V[g]V) relative to non-target VCV syllables by period ( $\pm 1$ SE).	37
2.3	Mean proportion of target CV syllables (e.g. [h]v or [g]v) relative to total CV syllables by period ( $\pm 1$ SE).	38
3.1	Preferential looking paradigm (similar to that used in Goldstein et al., 2010b).	57
3.2	Novel objects.	57
3.3	Mean proportion of looking to the target on Trial 1 by comprehension period and gender ( $\pm 1$ SE).	64
4.1	Playback paradigm.	91
4.2	Mean proportion of responses by Activity ( $\pm 1$ SE).	96
4.3	Mean proportion of responses to infant vocalizations by Activity and Response Type ( $\pm 1$ SE).	97
4.4	Relationship between response types during the play session and playback.	98
4.5	Mean ‘speechiness’ rating for each infraphonological type by parity group ( $\pm 1$ SE).	102
4.6	Mean proportion of responses to vowel stimuli by directedness of the vocalization ( $\pm 1$ SE).	104
4.7	Mean proportion of responses to vowel by vocal resonance of the vocalization ( $\pm 1$ SE).	105
4.8	Mean proportion of responses to vowel stimuli by directedness of the vocalization and response type ( $\pm 1$ SE).	106
4.9	Mean latency to produce a vocal response to vowel stimuli by directedness ( $\pm 1$ SE).	107
4.10	Mean proportion of responses to consonant-vowel stimuli by directedness of the vocalization and consonant-vowel type ( $\pm 1$ SE).	108
4.11	Mean proportion of responses to consonant-vowel stimuli by directedness of the vocalization and response type ( $\pm 1$ SE).	109
4.12	Mean proportion of responses to consonant-vowel stimuli by consonant-vowel type and response type ( $\pm 1$ SE).	110
4.13	Mean proportion of imitative responses to consonant-vowel stimuli by parity group ( $\pm 1$ SE).	111
4.14	Mean proportion of responses to phonological types by directedness of the vocalization and phonology type ( $\pm 1$ SE).	113
4.15	Mean latency of primiparous mothers to produce a vocal response by directedness and phonological type ( $\pm 1$ SE).	114
4.16	Mean proximity response by vowel type ( $\pm 1$ SE).	115
4.17	Mean proximity response to phonological types by directedness of the vocalization and phonology type ( $\pm 1$ SE).	117

## LIST OF TABLES

Table #	Table Caption	Page #
4.1	Stimulus Categories	93
4.2	Maternal Response Categories	95

## CHAPTER 1

### INTRODUCTION

Social interaction typically takes place through conversation in which the speaker and listener regularly switch roles by responding to each other's messages. While adult dialogue appears automatic and effortless, communication involves the rapid coordination of many cognitive functions. Communicators must convert incoming sensory input such as sound and nonverbal cues into meaningful linguistic messages, quickly interpret the message and then create a coherent motor response that the listener will comprehend. Garrod and Pickering (2004) suggest that efficient communication is possible because participants automatically undergo interactive alignment, a process in which communicators match their representations at multiple levels to establish common ground. For example, speakers mirror each other's body language, mannerisms, linguistic conventions, and even syntactic forms (Pickering & Garrod, 2006). Recent work suggests that alignment also takes place at the neural level through brain-to-brain coupling. The speaker and listener's neural activation in prefrontal, auditory, and visual cortices mirror each other with a slight temporal delay, suggesting that the perceptual system of one communicator aligns with the motor system of the other as a result of sharing information in a social environment (Hasson et al., 2012).

Just as conversation between adults involves bi-directional interaction between two communicators so too does caregiver-infant interaction. Infants are often thought of as passive receivers of information. However, infants actively contribute to communicative interactions by producing prelinguistic vocalizations. Around three months of age, repetitive and rhythmic patterns of interactions emerge through coordinated gaze, touch, and vocal communication

(Feldman, 2007). This early alignment is most evident in vocal turn-taking, in which caregivers initially organize vocal interactions by coordinating their responses to alternate with infants' vocalizations (Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001). Around 5 months of age, infants learn that their vocalizations elicit responses from caregivers (Goldstein, Schwade, & Bornstein, 2009). In the second half of the first year, caregiver-infant interactions become aligned around objects of joint attentional focus. Caregivers' responses to these triadic interactions with objects are related to later language development (Goldstein & Schwade, 2010; Tomasello & Farrar, 1986).

Once infants begin producing their first words, caregivers frequently cluster their utterances into partially overlapping sentences called variation sets (Waterfall, 2006). Variation sets are partial repetitions of speech grouped together in time (Küntay & Slobin, 1996; Waterfall, 2006). For example, a mother might say to her child, "Roll me the ball. Roll it. You rolled the ball to mama!" This type of alignment has been shown to facilitate toddlers' learning of novel verbs (Albert, Waterfall, Schwade, & Goldstein, under review). As young children begin producing multi-word utterances, caregivers continue to elaborate on children's responses by expanding their short utterances into full sentences. Caregivers' elaborations on children's speech are also positively correlated with later language outcomes (Newport, Gleitman & Gleitman, 1977; Hoff-Ginsberg, 1996).

Communicative alignment undergoes dramatic change as a result of the developing complexity of caregiver-infant interactions. By examining language development as an outcome of a socially embedded system of bidirectional feedback, the research presented in this dissertation investigates how prelinguistic infants and their caregivers develop vocal alignment to support the emergence of mature communication. The findings demonstrate that both infant

and caregiver behavior are tightly coupled and infant learning results from the alignment of caregiver and infant attention. Exploring the development of interaction dynamics between the caregiver-infant dyad is essential for understanding the mechanisms that support infant learning.

### **Historical Overview of the Function of Prelinguistic Vocalizations**

Traditionally, babbling was thought to be a form of motor practice, with no function in the development of communication and language (Kent, 1981). This perspective developed in the mid-20<sup>th</sup> century as a result of several attempts to describe the development of prelinguistic vocalizations using phonetic transcriptions based on the International Phonetic Alphabet (IPA; e.g. Jakobson, 1947 as cited by Oller, 2000; Irwin, 1947; Lenneberg, Rebelsky, & Nichols, 1965). In these early transcription studies, researchers documented only the infant vocalizations that fit the adult-based categories of the IPA while ignoring any vocalizations that did not fit into existing IPA categories. Consequently, infants' vocal repertoires were vastly misrepresented.

Nonetheless, four conclusions were formed based on these initial transcriptions of infant vocalizations that propagated through the literature on child development for decades (Oller, 2000). First, Jakobson (1947) concluded that infants produce all of the sounds of all of the worlds' languages, regardless of their caregivers' native language. Such a broad claim arose by forcing the messy syllables that infants produce into the categories of the IPA that are used to describe well-formed adult speech. Using IPA transcriptions, infants appeared to first make a wide range of uncommon syllables, followed by a 'silent period' where infants appeared to stop babbling for several months, after which a sudden transition occurred when infants began producing a more limited set of vocalizations that fit a mature consonant-vowel pattern of articulation. The early transcription work concluded that a discontinuity existed between babbling and production of first words. Infants were also presumed to babble at random, with no



influences from the environment. Specifically, Lenneberg (1967) and later Kent (1981) suggested that babbling was a byproduct of an immature motor system that developed without social feedback. Finally, after considering the vocalizations of a single deaf infant, it was presumed that deaf infants babble in the same ways as hearing infants (Lennenberg, Rebelsky, & Nichols, 1965). This result further reinforced the idea that input from the environment is irrelevant for vocal production. Taken together, several decades of research propagated the idea that babbling was simply a form of random motor activity that had no bearing on later language development.

The long-standing assumptions about the irrelevance of babbling to communication and language began to be questioned by studies that examined the development of babbling in more acoustic and behavioral detail, without relying on IPA transcriptions. Infants vocalize continuously throughout the first year (Vihman & Miller, 1988) thus refuting the idea of a discontinuity between babbling and early language. Over the course of the first year, the proportion of infant vocalizations containing consonants increases (Holmgren, Lindblom, Aurelius, Jalling, & Zetterstrom, 1986) and infants demonstrate increased variability in the range of consonants produced (Stoel-Gammon, 1988). In the second half of the first year, vocalizations drift towards the phonological patterns of their ambient language. Specifically, infants begin producing more vocalizations that contain common vowels (Boysson-Bardies et al., 1989), consonants (Boysson-Bardies & Vihman, 1991), and disyllable patterns (e.g. CVCV) that match their ambient language (Boysson-Bardies, 1993).

After recognizing the limitations of classifying infant vocalizations using the IPA, Oller took a taxonomic approach to vocal development and developed an infraphonological coding system based on both perceptual and acoustic features of speech, such as vowel resonance and

timing of consonant-vowel transitions. These features were used to define four infraphonological categories that describe the development of prelinguistic vocalizations (Oller & Lynch, 1992; Oller, 2000). Infraphonology refers to the infrastructure of a well-formed syllable. In the first two months of life, infants are in the phonation stage during which they produce *quasi-resonant* vowels. Quasi-resonant vowels are produced with a closed vocal tract (e.g. nasal vocalizations and grunts). Between 1-4 months of age, infants enter the primitive articulation stage and begin producing *fully-resonant* vowels. Fully-resonant vowels are produced with an open vocal tract. During the expansion stage, which begins between 3-8 months, infants start producing *marginal syllables*. Marginal syllables consist of slow sequences of consonant-vowel articulation with long transitions between consonant and vowel. Finally, between 5-10 months of age, infants enter the canonical stage and begin producing well-formed syllables. These *canonical syllables* are fully-resonant vowel nuclei combined with faster consonant-vowel transitions (e.g. [ba], [da]). The infraphonological coding system facilitates the study of infant vocal development by allowing all prelinguistic vocalizations to be categorized using the same definitions regardless of infant age.

Once infant vocalizations were described using the infraphonological coding system, it became apparent that infant's prelinguistic vocalizations undergo dramatic, experience-driven developmental change. Differences between the vocalizations of deaf and hearing infants became apparent when compared using an infraphonological coding system (Oller et al., 1985; Oller & Eilers, 1988). Particularly, deaf infants show a delayed onset of canonical babbling in comparison to hearing infants (Eilers & Oller, 1994). By taking a taxonomic approach to infant vocal development, these results demonstrate that babbling has functional significance as a precursor for later language. Infants do not randomly vocalize all of the sounds from every

language, but gradually develop stable patterns of vocalizing that match the language in the ambient environment.

The shift from immature prelinguistic vocalizations to well-formed syllables and first words is influenced by physiological and social development. In the first five months, infant vocal production is limited by physiological constraints of the vocal tract (Kent, 1981). As the vocal organs mature, infants begin producing rhythmic jaw oscillations (Davis & MacNeilage, 1995), which facilitates production of consonant-vowel syllables. While motor maturation certainly influences vocal development, social input is also necessary for normal speech development. Babbling develops through the interaction of feedback from the social environment coupled with increasing motor control over vocal organs. In the following sections, the influences of social feedback from the caregiver will be considered as well as the function of infants' prelinguistic vocalizations for creating a preparatory state that is conducive for learning.

### **Caregiver influences on language development**

Both the quality and quantity of maternal responsiveness to infant behavior is positively related to later language outcomes (Bornstein & Tamis-LeMonda, 1989; Rollins, 2003; Tamis-LeMonda, Bornstein, & Baumwell, 2001). The amount of speech infants hear in their first few years of life significantly impacts their vocabulary size and language abilities (Hart & Risley, 1995). Caregivers' descriptions and imitative responses to 9-month-old infants' vocalizations are positively correlated with the production of first words and multi-word utterances (Tamis-LeMonda, Bornstein, & Baumwell, 2001). The syntactic frames that caregivers use also predict productive vocabulary onset and other language milestones (Tomasello, 1995; Waterfall, 2006).

However, not all caregiver response types have the same impacts on later language development. Specific types of caregiver responses such as providing object labels (Stevens, Blake, Vitale & MacDonald, 1998) or asking questions (Furrow et al., 1979; Gleitman, Newport, & Gleitman, 1984) have positive relations with later language development. In contrast, redirective responses, such as labeling an object the infant is not attending to, negatively impact later vocabulary (Akhtar, Dunham & Dunham, 1991; Della Corte, Benedict & Klein, 1983; Tomasello & Farrar, 1986).

Caregivers' responses to the directedness of infants' vocalizations also have specific influences on later vocabulary. Infants tend to aim their vocalizations in one of three ways: caregiver-directed, object-directed, and undirected. Caregiver-directed vocalizations are those produced while the infant is looking at the mother's face. At 5 months of age, mothers' responses to infants' caregiver-directed vocalizations are positively correlated with later vocabulary at 18 months (Goldstein, Schwade, & Kirkpatrick, 2013). Object-directed vocalizations are defined as those produced while the infant is looking at an object that is held or within reach. Parental responsiveness to their 9-month-old infants' ODVs predicted infants' language development at 15 months (Goldstein & Schwade, 2010). Undirected vocalizations are vocalizations produced at neither an object nor a caregiver and can be characterized as vocalizing into empty space.

The studies above are all correlational in design and examine the effects of parent-infant interaction over long time periods. As a result, caregiver interactions are recognized as an important factor in infant language development but much less is known about the proximal patterns of caregiver responding that may facilitate language learning in real-time. Studying the spatial and temporal structure of parental behavior and infant learning as they co-occur in real

time will shed light on the mechanisms responsible for the earliest stages of language development.

In addition, when investigating the influences of caregiver responsiveness to infant vocalizations, past research categorized caregivers' responses to prelinguistic vocalizations without regard to the acoustic qualities or contexts in which the vocalization was produced. However, as discussed in the previous section, infants make a wide range of vocalizations of varying qualities and can direct those vocalizations in different ways. Previous work suggests that caregivers adjust their responses based on the qualities of infants' vocalizations (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008) and tend to respond to more mature vocalizations (Gros-Louis et al., 2006), but these effects have yet to be experimentally manipulated. The experiments presented in this dissertation investigate the social feedback system between caregivers and infants to systematically titrate out (i.e. "isolate and test a number of elements that might be important..."; Alberts & May, 1984, 162) the properties of infant behavior that influence caregiver responding.

A second understudied question is whether caregiving behavior develops and if so how? In the parenting literature, parents are rarely compared to non-parents (but see Green, Jones & Gustafson, 1987 for comparison of parents' versus non-parents' perceptions of infant cries). Without a baseline for comparison, it is unknown whether parents respond differently to infants' prelinguistic vocalizations than non-parents. However, using nulliparous females as a baseline in which to compare experienced females against is a standard practice in the non-human animal literature. For example, nulliparous juvenile females provide non-vocal reinforcement in the form of wing strokes indiscriminately to male song but experienced females are more selective in their responding and only provide wing strokes to well-formed elements of birdsong (West et al.,

2006). My dissertation incorporates insights and methods from the non-human animal literature to investigate how women of differing parity respond to infant vocalizations and how the form and timing of parental behavior relates to infant learning.

### **Influences of Proximal Social Interactions on Language Development**

The ability of infants to communicate vocally undergoes remarkable change over the first year. From the relatively fixed signals of crying, through many months of variable and immature vocalizing, to the first words, the communicative characteristics of vocalizations gradually assume the flexible and meaningful functions we associate with mature speech and language. What are the underlying mechanisms that facilitate this shift to more mature vocalizations? How do infants determine which patterns of sounds are relevant to learn? My dissertation takes an ecological approach to link infant learning and parental behavior within the same system. An ecological approach to development states that the environment contains many affordances, or opportunities for action that immature learners can utilize (Gibson & Pick, 2000). In the case of infant learning, caregivers organize the environment in many ways that are helpful for learning and infants learn the structure that exists in caregiver behavior.

Infants readily detect regularities and extract structure from sequences of sensory input. When exposed to a continuous stream of speech, 8-month-old infants focus on the structure of the input to detect distributional probabilities in the patterns and recognized violations of those patterns with just a few minutes of exposure (Saffran, Aslin, & Newport, 1996). This process of statistical learning in perception has been robustly demonstrated in both auditory and visual modalities with a range of stimuli including speech, shapes, and visual scenes (Fiser & Aslin, 2002; Gomez, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996).

Infants quickly use the information gained via statistical learning and apply it to new situations. For example, infants immediately associated the patterns of phonemes they parse with novel objects (Graf-Estes et al., 2007) and the transitional probabilities extracted early in life influence the types of words infants will learn months later (Graf-Estes & Bowen, 2013). With exposure to speech, infants develop categories that match the phonological contrasts present in the ambient language (Best, 1994; Jusczyk, 1992). However, speech perception is only one half of language development. While infants can demonstrate statistical learning rapidly in laboratory tasks when passively exposed to stimuli, how do they translate knowledge of relevant phonological forms into their babbling?

Many structural regularities exist in infants' environments that provide opportunities for learning. Caregivers' behavior is organized in predictable ways, providing opportunities for infants to detect and evaluate patterns (Goldstein et al., 2010). For example, caregivers' speech to infants differs from speech to adults in that it is slower, contains longer pauses, simplified sentence structure and contains higher and more exaggerated pitch contours (Englund & Behne, 2006; Kuhl, 2007). Caregivers tend to coordinate their actions on objects with infant-directed speech, creating intersensory redundancy (Brand & Baldwin, 2002; Koterba & Iverson, 2009; Gogate, Bahrick, & Watson, 2000). This motion-speech synchrony creates structure between speech and action, which increases the saliency of the object being acted upon. Caregivers often speak to infants in very short sentences and tend to repeat the same sentence frames (Cameron-Faulkner, Lieven & Tomasello, 2003). Hearing nouns in these simplified sentence frames facilitates word recognition (Fernald & Hurtado, 2006). Thus, infant-directed speech reduces the complexity of the infant's environment and highlights the structural regularities in caregiver behavior.

In addition to organizing infant attention with IDS, caregivers also frequently provide prompt, contingent feedback to prelinguistic vocalizations (Goldstein & West, 1999). As a result, infants quickly learn that their vocalizations elicit caregiver responses. By 5 months of age, infants associate their vocalizations with contingent responses from caregivers (Goldstein, Schwade, & Bornstein, 2009). By 9 months of age, infants increase their vocal production and modify their vocalizations to be more speech-like in response to caregivers' contingent input (Goldstein & Schwade, 2008). For example, infants rapidly increase their production of consonant-vowel syllables when caregivers contingently reply to infant vocalizations by speaking consonant-vowel (CV) syllables (Goldstein & Schwade, 2008). However, infants non-contingently exposed to the same input do not modify their vocal production.

Infants also learn and reproduce new phonological patterns when phonological input is presented contingently following babbling and contains phonemic variability that highlights the underlying phonological regularities (Goldstein, Syal, & Schwade, under review). English-speaking infants who were contingently exposed to a relatively rare phonological pattern (VCV) discovered the statistical regularities in their mothers' speech and began producing significantly more VCV utterances. However, infants did not reproduce the surface characteristics of specific phonemes (e.g. [aba] versus [adi]) that their mothers were cued to use when responding. Contingent social input thus provides information about the structural regularities of language and promotes learning. Further, infants who received contingent but repetitive input did not learn the rare phonological pattern, demonstrating that phonemic variability is necessary for infants to extract the underlying phonological pattern.

Caregivers' contingent and variable responses to infant babbling provide reliable cues that aid infants in learning to produce the phonological patterns of their ambient language. As



infants vocalize, caregivers promptly respond, and infants then modify their vocal production. Such a bidirectional system of vocal feedback facilitates the development of speech and early language and is evidence of vocal alignment (Goldstein & Schwade, 2008). This process of extracting regularities from caregiver's patterned input is known as *socially guided statistical learning* (SGSL) and has been demonstrated to be a mechanism underlying vocal development in infants (Goldstein & Schwade, 2008; 2010; Goldstein et al., 2010). Only by investigating the caregiver-infant dyad as a developing system does it become apparent that infant vocal development is happening as a result of active interaction by both the infant and caregiver.

## **Summary of the Dissertation**

My dissertation will investigate the role of alignment and SGSL as mechanisms underlying early language development and communication by examining the bidirectional influences that exist in the infant-caregiver dyad. Specifically, my dissertation research will titrate out the critical cues used by infants and caregivers to establish and maintain vocal alignment. My research objectives are to: (1) investigate possible motoric constraints on socially guided statistical learning for infant's vocal production; (2) determine the function of the infant's own object-directed prelinguistic vocalizations in preparing the infant to learn and facilitating the mapping of word-object associations; (3) examine the role of caregiver experience in responding to infant vocalizations; and (4) examine the characteristics of infant vocalizations that elicit responses from caregivers.

Infants rapidly learn new vocal patterns from contingent social feedback to their babbling (Goldstein & Schwade, 2008; Goldstein, Schwade, & Syal, under review). A relatively

understudied question is what are the constraints on SGSL for vocal learning in infants? In other words, what exactly is being learned when infants produce new phonological using socially guided statistical learning? Some have posited that early babbling may be constrained by motor articulation (MacNeilage, 2008). However, others have argued that infants' early vocal production is flexible and guided by feedback from the environment (Oller, 2000). In Chapter 2, I test whether infants learn new phonological regularities in terms of a specific oscillatory motor pattern that is subject to inherent constraints or whether infants learn the acoustic structure of the phonological pattern itself.

In Chapter 3, I investigate the function of infant babbling for preparing the infant for learning. Recent research in our lab has determined that babbling plays a crucial role in learning early words. Specifically, word learning is facilitated by contingent caregiver responses to infants' *object-directed vocalizations (ODVs)*. ODVs are defined as vocalizations made when an infant is looking at an object that is within reach or being held. Eleven-month-old infants are more likely to learn the name for an object if it is labeled just after an ODV rather than after a look alone (Goldstein et al., 2010). What mechanism is responsible for the facilitative effects of ODVs on word learning? I hypothesized that the act of babbling gates learning by increasing arousal as a form of vocal self-stimulation. Vocal self-stimulation has been found to induce a state of anticipatory readiness in several species (Panksepp, 1998). In the case of infant learning, I predicted that when infants babble, the stimulation caused by vocalizing induces a state of anticipatory readiness that helps to organize attention and thereby facilitate learning. To test this hypothesis, I attempted to induce a state of readiness by playing infants their own prerecorded ODV over a speaker prior to labeling. Infants who heard their own vocalization prior to labeling learned as well as the infants in the previously published study who heard the label after

producing a spontaneous ODV. Thus, when infants produce ODVs, they may actually be in a state that is conducive to learning, more so than at other times. If infants' vocalizations prepare infants to learn, their state of anticipatory readiness provides an explanation as to why caregivers' responses to infant vocalizations have such strong effects on later learning.

Finally, most previous work on the effects of socially guided learning on prelinguistic vocal development has manipulated caregivers' responses to their infants' sounds (e.g. Chapter 2; Goldstein & Schwade, 2008). The results of Chapter 2 demonstrate that caregivers' responses provide reliable cues that aid infants in learning the patterns of production that characterize the ambient language. However, the characteristics of infant vocalizations that drive caregiver responses are unknown. In Chapter 4, I systematically manipulate infants' vocalizations to examine their effects on the responses of caregivers using a playback paradigm.

Playback paradigms are widely used in the field of animal communication to assess the potency of a vocal signal on the receiver. In a playback paradigm, a subject is exposed to prerecorded stimuli of a conspecific and the subject's behavioral response to each stimulus is measured. The playback experiments presented here are the first to explore how acoustic characteristics of infants' vocalizations influence caregiver responsiveness. Caregivers predicted their own behavior in response to highly controlled audiovisual stimuli. These final experiments had three goals: (1) validate the playback paradigm as a measure of caregiver responsiveness; (2) determine the acoustic and contextual parameters of babbling that influence caregiver responses; and (3) assess the role of experience on caregiver responses to prelinguistic vocalizations. Taken together, the experiments presented in this thesis take a systems approach to development, informed by ecological and comparative perspectives, to enhance our understanding of the

mechanisms by which interactive alignment and socially guided statistical learning influence vocal development and language learning.

## References

- Akhtar, N., Dunham, F. & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, 18, 41-49.
- Albert, R. R., Schwade, J. A., Waterfall, H. R., & Goldstein, M. H. (under review). Looking, pointing, and demonstrating: The effects of variation in speech and exemplars on multiple measures of verb learning.
- Alberts, J.R. & May, B. (1984). Nonnutritive, thermotactile induction of filial huddling in rat pups. *Developmental Psychobiology*, 17, 161-181.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.
- Bornstein, M. H. & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *New Directions in Child Development*, 43, 49-61.
- Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In B. Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 353–363). Dordrecht: Kluwer.
- Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297–320.
- Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1–17.

- Brand, R.J. & Baldwin, D.A. (2002). Motherese. *Child Development, Volume One in the Macmillan Psychology Reference Series*, (pp. 278-279). Farmington Hills, MI: The Gale Group.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843-873.
- Davis, B. L. & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language and Hearing Research*, 38, 1199-1211.
- Della Corte, M., Benedict, H. & Klein, D. (1983). The relationship of pragmatic dimensions of mothers' speech to the referential-expressive distinction. *Journal of Child Language*, 10, 35-43.
- Eilers, R. E. & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, 124, 199-203.
- Englund, K. & Behne, D. (2006). Changes in infant directed speech in the first six months. *Infant and Child Development*, 15, 139-160.
- Feldman, R. (2007). Parent-infant synchrony and the construction of shared timing: Physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry*, 48, 329-354.
- Fernald, A. & Hurtado, N. (2006). Names in frames: infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9, 33-40.
- Fiser, J. and Aslin, R.N. (2002) Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822-15826.
- Furrow, D., Nelson, K. & Benedict, H. (1979). Mothers' speech to children and syntactic development: Some simple relationships. *Journal of Child Language*, 6, 423-442.

- Garrod, S. & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8, 8–11.
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. Oxford: Oxford University Press.
- Gleitman, L. R., Newport, E. L. & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, 11, 43-79.
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71, 878–894.
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, 113, 52–58.
- Goldstein, M.H. & Schwade, J.A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *The Oxford Handbook of Developmental and Comparative Neuroscience*. Oxford University Press.
- Goldstein, M.H. & Schwade, J.A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515-523.
- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants association their own noncry vocalizations with responses from caregivers. *Child Development*, 80, 636-644.
- Goldstein, M.H., Schwade, J.A., & Kirkpatrick, F. (2013, April). *Maternal responsiveness to socially-directed prelinguistic vocalizations predicts later language development*. Paper

presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Goldstein, M.H., Syal, S., & Schwade, J.A. (submitted). *Socially guided statistical learning drives the emergence of speech in infants*. Manuscript submitted for publication.

Goldstein, M.H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, 15, 362-391.

Goldstein, M.H., Waterfall, H.R., Lotem, A., Halpern, J.Y., Schwade, J.A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and place. *Trends in Cognitive Sciences*, 14, 249-258.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.

Graf-Estes, K. & Bowen, S. (2012). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology*, 114, 405-417.

Graf-Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18, 254–260.

Green, J.A., Jones, L.E., & Gustafson, G.E. (1987). Perception of cries by parents and nonparents: Relation to cry acoustics. *Developmental Psychology*, 23, 370-382.

Gros-Louis, J., West, M.J., Goldstein, M.H., & King, A.P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development* 30, 112-119.



- Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Brookes: Baltimore, MD.
- Hasson, U. Ghazanfar, A.A, Glantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16, 114-121.
- Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, 22, 155–163.
- Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterstrom, R. (1986). On the phonetics of infant vocalization. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 51-63). New York: Stockton Press.
- Irwin, O.C. (1947). Infant speech: Consonant sounds according to manner of articulation. *Journal of Speech and Hearing Disorders*, 12, 402-404.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy. *Monographs of the Society for Research in Child Development*, 66, i-149.
- Juszyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17–64). Timonium, MD: York Press.
- Kent, R. D. (1981). Articulatory-acoustic perspectives on speech development. In R. E. Stark (Ed.), *Language behavior and early childhood* (pp. 105-126). New York: Elsevier.
- Kirkham, N Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidences for a domain general learning mechanism. *Cognition*, 83, 835-842.
- Koterba, E. A. & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development*,

32, 437- 444.

Kuhl, P.K. (2007). Cracking the speech code: how infants learn language. *Acoustic Science and Technology*, 28, 71-83.

Küntay, A., & Slobin, D. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 265–286). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Lenneberg, E., Rebelsky, F. G., & Nichols, I. A. (1965). The vocalizations of infants born to deaf and hearing parents. *Human Development*, 8, 23-37.

MacNeilage, P. F. (2008). *The origin of speech*. Oxford: Oxford University Press.

Newport, E. L., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–150). Cambridge: Cambridge University Press.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of vocal development and disorders. In C. A. Ferguson, L. Menn & C. Stoel- Gammon (Eds.), *Phonological development* (pp. 509–536). Timonium, MD: York Press.

Oller, D. K. & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59, 441-449.

- Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech and Hearing Research*, 28, 47–63.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.
- Pickering, M.J. & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4, 203-228.
- Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, 24, 221-234.
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Stevens, E., Blake, J., Vitale, G., & MacDonald, S. (1998). Mother-infant object involvement at 9 and 15 months: Relation to infant cognition and early vocabulary. *First language*, 18, 203- 222.
- Stoel-Gammon, C. (1988). Prelinguistic vocalizations of hearing-impaired and normally hearing subjects: a comparison of consonantal inventories. *The Journal of Speech and Hearing Disorders*, 53, 302-315.
- Tamis-LeMonda, C.S., Bornstein, M.H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72, 748-767.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 53-71.
- Tomasello, M. (1995). *Constructing a language*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57, 1454-1463.

Vihman, M. M., & Miller, R. (1988). Words and babble at the threshold of language acquisition. In M. D. Smith & J. L. Locke (Eds.), *The emergent lexicon: The child's development of a linguistic vocabulary* (pp. 151–183). San Diego: Academic Press.

Waterfall, H.R. (2006). *A little change is a good thing: Feature theory; language acquisition and variation sets*. Ph.D. thesis, University of Chicago.

West, M. J., King, A. P., White, D. J., Gros-Louis, J., & Freed-Brown, G. (2006). The development of local song preferences in female cowbirds (*Molothrus ater*): Flock living stimulates learning. *Ethology*, 112, 1095-1107.

## CHAPTER 2

THE ‘AHA’ MOMENT: WHAT IS LEARNED WHEN INFANTS LEARN  
TO PRODUCE NEW PHONOLOGICAL REGULARITIES?

## **Abstract**

Infants learn the phonological patterns of their ambient language by discovering the statistical regularities present in caregivers' contingent speech following infant babbling. While infants use this feedback from caregivers to guide their vocal production, it is unclear at what level infants learn phonological regularities. Infants could learn phonology in terms of both the motor patterns of jaw oscillations needed to produce the pattern and in terms of the acoustic patterns present in the phonological input. However, the relative influences of the motor system on vocal learning is unknown because in previous vocal learning studies, caregivers responded to their infants' vocalizations with phonological patterns that could be learned as either a motor pattern or acoustic pattern. Here, caregivers were instructed to contingently respond to their 9-month-old infant's vocalizations with a rare phonological pattern that did not require a jaw oscillation to produce. Thus, if infants were to learn the pattern, they have learned the phonological regularities as an acoustic pattern. Infants increased their production of vocalizations that matched the rare phonological pattern demonstrating that, at least by 9-months of age, infants learn phonology at the level of an acoustic pattern.

## Introduction

Over the first year of life, infants' prelinguistic vocalizations come to resemble the phonological patterns of the ambient language (Boysson-Bardies, 1993; Boysson-Bardies & Vihman, 1991). Such stable phonological patterns of babbling emerge as infants learn from the distribution of phonemic and phonological features that characterize their language (Vihman & Miller, 1988; Vihman, 1993). What mechanisms drive the process of phonological learning? Caregivers' responses to infants' prelinguistic vocalizations provide reliable cues that aid infants in learning the patterns of production that characterize the ambient language. In response to caregivers' contingent feedback, infants increase their vocal production and modify their vocalizations based on the phonological regularities of caregivers' speech (Goldstein & Schwade, 2008; Goldstein, Syal, & Schwade, under review). This process is known as *socially guided statistical learning* (SGSL) and is demonstrated to be a mechanism underlying vocal development in infants (Goldstein & Schwade, 2008; 2010; Goldstein et al., 2010).

However, it is unclear what exactly is learned from phonological input and at what level learning occurs. Are infants learning to produce specific patterns of sounds, or are they learning to coordinate their articulators in specific ways, or both? Several theories have been proposed with regards to what constraints might guide infant vocal development. Some argue that physiology constrains articulatory movement, limiting vocal development, and in turn the types of vocal patterns infants are capable of learning and producing (MacNeilage, 1998; 2008). In this view, the distribution of phonological features is learned at the level of motor patterns. For example, certain consonant-vowel (CV) sequences tend to be commonly produced and repeated by infants (MacNeilage & Davis, 1990). According to the Frame/Content theory of speech production (MacNeilage, 1998; 2008), the physiological constraints of mandibular oscillation

constrain infants' early vocalizations via a dominant frame of jaw oscillation (open-closed pattern) to produce stable patterns of babbling. The Frame/Content theory further posits that infants do not yet have control of the specific components of CV articulation, but produce variable babbling patterns as a result of random movement of the tongue and lips within the CV frame (MacNeilage & Davis, 2000). Only as infants gain experience with articulatory movements and are exposed to specific syllables do the content elements (specific vowels and consonants) become differentiated from the frame. In this view, infants should have more control over the starting position of the oscillation frame (i.e. closed for [ba] or open for [aba]) than the specific sounds that result from their vocalizations. Thus, if infants are utilizing socially guided statistical learning to pick up new phonological patterns, it must be within the constraints of their dominant oscillation pattern.

A contrasting perspective argues that infants' early vocalizations are chiefly characterized by contextual freedom and functional flexibility (Oller et al., 2013). Unlike the relatively fixed signals of crying and laughing, infants' vocalizations convey a range of emotional information depending on the context in which they are produced. Such functional flexibility suggests that vocal development is less constrained (Oller, 2000) and phonological patterns are learned as a series of auditory patterns over and above motor constraints. In this view, infants' speech becomes canalized as a result of rhythmic jaw movement, but auditory experience is the driving force behind vocal development. For example, the babbling patterns of deaf infants are different from that of hearing infants in that deaf infants show delays in the onset of producing well-formed canonical syllables (Eilers & Oller, 1994). Further, infants who are tracheostomized in the first two years of life, and thus unable to practice phonating, rapidly produce their first words after only a short period following decannulation (Locke & Pearson, 1990; Ross, 1983).



Auditory experience is therefore an important mediator of motor movements when infant vocalize and appears to be sufficient for learning phonological patterns.

The present study tests these two theories to assess the relative influence of articulatory constraints on vocal learning. While MacNeilage (2008) and Oller (2000) both make theoretical claims about vocal development in human infants, these theories are yet to be tested with infants in a learning environment. Studies of vocal learning in other species, however, offer clues for understanding relevant mechanisms in humans. Songbirds have long been considered a model for human speech and early language learning (Doupe & Kuhl, 1999; Goldstein et al., 2003). A longstanding question in the songbird literature was whether birdsong is learned as a series of vocal motor gestures or as a sequence of acoustic units. In other words, do birds learn song in terms of the motor patterns or auditory patterns necessary to produce song?

Traditionally it was believed that song in species such as zebra finch was produced as a series of motor gestures, independent of sensory feedback (Price, 1979). However, manipulations such as deafening (Brainard & Doupe, 2000) or temporarily muting birds (Pytte & Suthers, 1999) revealed that auditory feedback is necessary for birds to learn species-typical song. Songbirds such as zebra finch and cardinals appear to learn song in terms of acoustic patterns. Recent work suggests that birds develop an acoustic template while listening to the song of conspecifics and then use that template to refine their motor gestures (Mendez, Dall'Asen, Cooper, & Goller, 2010). For example, cardinals adjust their patterns of articulatory production to continue to produce species-typical song when their vocal muscles are altered (Suthers, Goller, & Wild, 2002). Motor production of birdsong is therefore mediated by auditory feedback. Thus, birds learn the songs of conspecifics as a series of acoustic patterns, which then guides their motor production.

Based on the songbird work, we predict that infants will also be capable of learning the regularities of speech in terms of acoustic patterns. However, previous studies of socially guided vocal learning in infants have focused on easy-to-pronounce vowel, consonant-vowel, or vowel-consonant-vowel patterns that did not discriminate between motor and acoustic learning. These studies revealed rich social support and guidance for production of phonological patterns. Caregivers are consistently responsive to infants' vocalizations (Goldstein & West, 1999), providing infants with structured feedback that highlights the distributional regularities of the ambient language. By 9 months of age, infants increase their vocal production and modify their vocalizations to be more speech-like in response to their caregiver's variable and contingent input (Goldstein & Schwade, 2008). Specifically, infants rapidly increase their production of consonant-vowel syllables when caregivers contingently reply to infant vocalizations by speaking consonant-vowel (CV) syllables (Goldstein & Schwade, 2008). Conversely, infants increase their production of fully-resonant vowels if caregivers contingently utter variable vowels in response to infant babbling. Infants non-contingently exposed to the same input do not modify their vocal production. Thus, stable phonological patterns of babbling emerge as infants learn from contingent social feedback that highlights the regularities of the ambient language.

Further, infants learn and reproduce new phonological patterns when they are presented contingently on babbling and contain phonemic variability that highlights underlying phonological regularities (Goldstein, Syal, & Schwade, under review). English-speaking infants who were contingently exposed to a relatively rare phonological pattern (vowel-consonant-vowel) discovered the statistical regularities in their mothers' speech and began producing significantly more VCV utterances. However, infants did not reproduce the surface characteristics of specific phonemes (e.g. [aba] versus [adi]) that their mothers used. Further,

infants who received contingent but repetitive input (identical VCV utterances) did not learn the non-native phonological pattern, demonstrating that phonemic variability is necessary for infants to extract the underlying phonological pattern. Taken together, these results demonstrate that caregivers' contingent and variable responses to infant babbling provide reliable cues that aid infants in learning to produce the phonological patterns of their ambient language.

A limitation of these vocal learning studies is that they used sound patterns that require a mandibular oscillation to produce. Therefore, the question of what exactly infants learn when they learn from social feedback still remains. If the frame/content theory of speech production is correct, infants in previous vocal learning studies (e.g. Goldstein, Syal, & Schwade, under review) may not have been learning the underlying phonological pattern, but rather mimicking mandible movements without regard for the sounds produced. For example, when producing a CV syllable such as [ba], the lips and jaw begin in a closed position and open to produce the vowel. When exposed to a VCV pattern, it may have been easier for infants to switch the starting position of the mouth from closed to open to produce the new pattern than it was for them to utter specific phonemes. In contrast, if phonological learning is not constrained by mandibular oscillation patterns and infants learn phonological patterns acoustically, we would expect that infants could acquire a new phonological pattern that does not require a jaw oscillation.

In the current experiment, we used a vocal learning paradigm to assess the relative influences of motor constraints and statistical learning on infants' abilities to learn new phonological patterns. To test whether SGS is constrained by mandibular oscillation as predicted by the frame-content theory, we presented infants with a phonological pattern that can be produced without mandibular oscillation. We tested infants' abilities to learn one of two new VCV forms, a vowel- [h]-vowel pattern and a vowel- [g]-vowel pattern. Infants commonly

produce both [h] and [g] syllables (Vihman, 1992). If infants are less constrained by a mandibular oscillation pattern and can learn phonology in terms of the distribution of sound patterns, then they should increase their production of the target input. However, if production is constrained by a mandibular oscillation pattern, then infants should not increase production of the target VCV pattern (V[h]Vs and V[g]Vs) but should substitute a consonant (e.g. V[b]V) that maintains the oscillatory frame. Either result would provide insight into the role of socially guided statistical learning for infant vocal learning.

## **Method**

### *Participants*

Sixty 9.5-month-old infants (mean age: 9 months, 24 days, range: 8 months, 9 days to 10 months 21 days) participated with a primary caregiver (n= 49 with mother). Thirty infants were assigned to the V[h]V input condition (n= 15 male) and 30 infants to the V[g]V input condition (n= 15 male). An additional 18 infants were tested but excluded because the infant cried or fussed excessively (n= 4), the caregiver failed to follow directions (n= 4), there was an equipment failure (n= 4), the dyad could not return for the second session (n= 1) or the infant produced fewer than five vocalizations per period (n= 5), which was below the 10<sup>th</sup> percentile for number of vocalizations produced per period. To recruit participants, letters and emails were sent to families listed in the local newspaper's birth announcements. For their participation, caregivers received an infant t-shirt, bib, or children's book.

### *Apparatus*

Infants were tested in a 3.7-m x 5.5-m playroom containing infant toys, a toy box, and pictures of animals on the walls. Infants could move around freely to explore; the size of the

room allowed infants to play without continuous interaction with their caregiver. Dyads were video recorded from one of three remote-controlled cameras mounted on the walls. To obtain accurate and detailed recordings of the infant's vocalizations, the infant wore a wireless microphone (Telex ELM-22; Telex Communications, Inc., Burnsville, MN) and transmitter (Telex USR-100) concealed inside the lining of adjustable denim overalls. To obtain accurate recordings of the caregiver's speech, caregivers also wore a wireless label microphone and transmitter in a pouch around the waist. During the second session (see procedure), the experimenter used a head microphone (Audio Technica Pro-8) to give the caregivers instructions via wireless headphones (2.4GHz Digital Wireless Stereo Headphones (Model 33-103; RadioShack). The infant's vocalizations were recorded on a separate audio channel from the caregiver and experimenter's speech.

### *Stimuli*

During the social response period, caregivers were cued to respond to their infants with VCV-patterned disyllables from a randomized list of 24 VCV disyllables. Infants were assigned to one of two conditions (V[h]V input, V[g]V input). Gender was balanced within condition. Caregivers with infants in the V[h]V input condition were cued to respond with V[h]V-patterned disyllables while caregivers with infants in the V[g]V input condition were cued to respond with V[g]V-patterned disyllables. The vowel position was filled by one of six vowels (ə, o, u, ε, i, a), that are typical of English (Handbook of the International Phonetic Association, 1999) and commonly produced by infants (Stoel-Gammon & Herrington, 1990). However, each VCV cue contained either only front/mid vowels (e.g., [aε]) or back/mid vowels (e.g., [ogu]) because a

VCV that contains both back and front vowels (e.g., [ago]) would require a mandibular oscillation to produce.

### *Procedure*

Caregivers brought their infants into the laboratory for two 30-min play sessions on consecutive days. The first visit was a familiarization session to allow infants and caregivers to get comfortable with the playroom. The second session followed an ABA design and was divided into three 10-min periods: Baseline-1, Social Response, and Baseline-2. During the baseline periods, caregivers were asked to play with their infants as they would at home. To investigate the effect of contingent VCV input on infant's vocal production, caregivers' responses to infants' vocalizations were manipulated during the social response period. At the beginning of the social response period, the experimenter instructed the caregiver over the headphones on how to react to the infant's vocalizations. The caregiver was told that each time the experimenter cued her with a VCV disyllable, she should lean in, smile, touch her baby and repeat the cued disyllable she heard. In the absence of instructions from the experimenter, she was to continue speaking and playing with her infant as before.

### *Data Coding and Analysis*

To calculate the frequency of infant vocalizations produced in each period, vocalizations were divided into syllables. Each syllable was comprised of a single vowel (V) or a consonant and a vowel (either CV or VC). To evaluate infants' production of the target consonant, CV and VC syllables containing an [h] (in the V[h]V condition) or a [g] (in the V[g]V condition) were categorized. To evaluate infants' learning of the new VCV pattern, infant vocalizations were additionally classified according to their phonological patterning. Vocalizations were categorized

as VCV, V[h]V, or V[g]V when a V was followed by a CV syllable without a breath between the two syllables. Fusses, raspberries and vegetative sounds (e.g. coughs) and sounds with oral obstructions (e.g. toys in the mouth) were excluded from analyses. The first author coded 100% of the data while one of three research assistants independently recoded 50% of the data (all 3 periods from 30 infants) to assess reliability. Mean reliability was  $r = .89$  (range= .75-1.0) for phonological patterning,  $r = .913$  for VCVs,  $r = .86$  for V[h]Vs,  $r = .84$  for V[g]Vs,  $r = .87$  for [h]v syllables, and  $r = .86$  for [g]v syllables. All disagreements were discussed and resolved; analyses were conducted on the resolved data.

To assess the effects of the target input on vocal production, three variables were computed for each of the three periods. *Target VCVs* were defined as VCV disyllables in which the consonant matched the consonant in the caregiver's VCV input (i.e. V[h]Vs for infants in the V[h]V input condition and V[g]Vs for infants in the V[g]V input condition). *Non-Target VCVs* were defined as all other VCV disyllables with non-target consonants. We calculated a proportion of Target VCVs relative to Non-Target VCVs. Finally, we calculated the proportion of CV syllables that contained the target consonant relative to the total number of CV syllables produced.

## Results

### *Frequency of VCV disyllables*

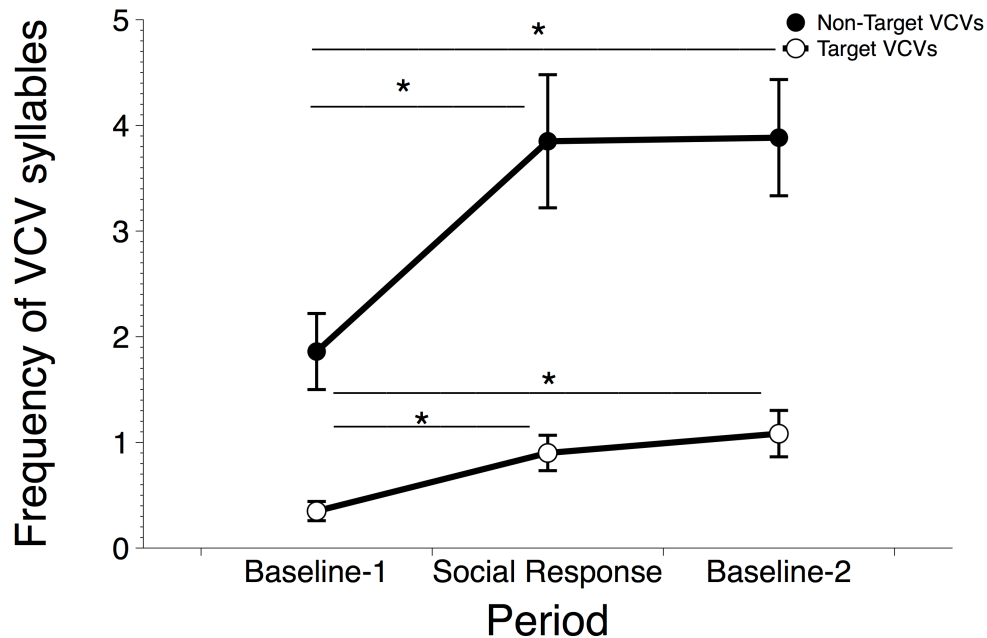
To assess changes in infant's vocal production patterns, a 3 (Period: Baseline-1, Social Response, Baseline-2) x 2 (Pattern: Target VCVs, Non-Target VCVs) x 2 (Condition: V[h]V input, V[g]V input) x 2 (Gender) mixed ANOVA was conducted on the frequency of VCV-patterned disyllables produced. There was a significant effect of Period,  $F(2, 112) = 11.889$ ,

$p < .001$ ,  $\eta_p^2 = .175$ . Infants increased their production of VCV disyllables from Baseline-1 to Social Response (Tukey's HSD  $p = .001$ ), and produced more VCV disyllables during Baseline-2 than Baseline-1 (Tukey's HSD  $p = .001$ ). There was also a significant main effect of Pattern,  $F(1, 56) = 37.308$ ,  $p < .001$ ,  $\eta_p^2 = .400$ . Infants produced more Non-Target VCVs than Target VCVs in all three periods. These main effects were qualified by a significant Period x Pattern interaction,  $F(2, 112) = 3.649$ ,  $p = .029$ ,  $\eta_p^2 = .061$  (Figure 2.1). The interaction was decomposed by Pattern.

For production of Target VCVs, there was a significant main effect of period,  $F(2, 112) = 6.468$ ,  $p = .002$ ,  $\eta_p^2 = .104$ . Infants increased their production of target VCV disyllables from Baseline-1 to Social Response, Tukey's HSD  $p = .001$ . Further, infants produced more target VCV disyllables during Baseline-2 than Baseline-1, Tukey's HSD  $p = .002$ .

For production of Non-target VCVs, there was also a significant main effect of period,  $F(2, 112) = 8.291$ ,  $p < .001$ ,  $\eta_p^2 = .129$ . Infants increased their production of non-target VCV disyllables from Baseline-1 to Social Response, Tukey's HSD  $p = .002$ . Further, infants produced more non-target VCV disyllables during Baseline-2 than Baseline-1, Tukey's HSD  $p = .001$ . There were no other significant main effects or interactions from the mixed ANOVA,  $ps > .096$ .





*Figure 2.1.* Mean frequency of target VCV disyllables (e.g. V[h]V or V[g]V) and non-target VCV disyllables by period ( $\pm 1$  SE). Infants increased their production of both target VCVs and non-target VCVs from baseline-1 to social response and maintained that increase during baseline-2. Infants produced more non-target VCVs than target VCVs in all three periods, which is not surprising because there are more possible non-target VCV combinations than there are target VCV combinations.

#### *Proportion of Target VCVs to Non-Target VCVs*

To assess changes in infant's target disyllable production relative to non-target disyllable production, a 3 (Period: Baseline-1, Social Response, Baseline-2)  $\times$  2 (Condition: [h], [g])  $\times$  2 (Gender) mixed ANOVA was conducted on the proportion of target VCV disyllables produced to non-target VCV disyllables produced. There was a trend towards a significant effect of Period,  $F(2, 112) = 2.657, p = .075, \eta_p^2 = .045$  (Figure 2.2). Infants tended to increase their proportion of target VCVs from baseline to social response. There were no other significant main effects or interactions,  $ps > .084$ .

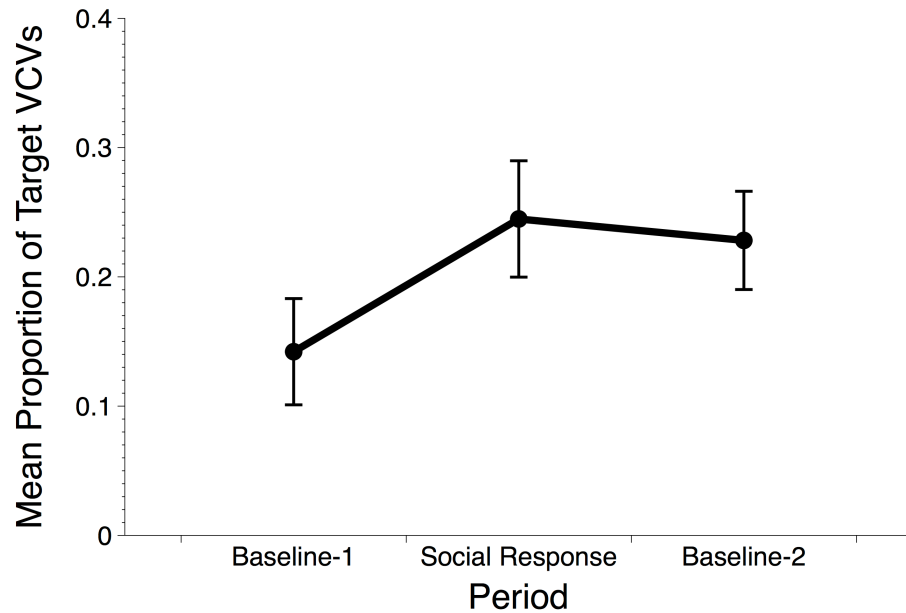
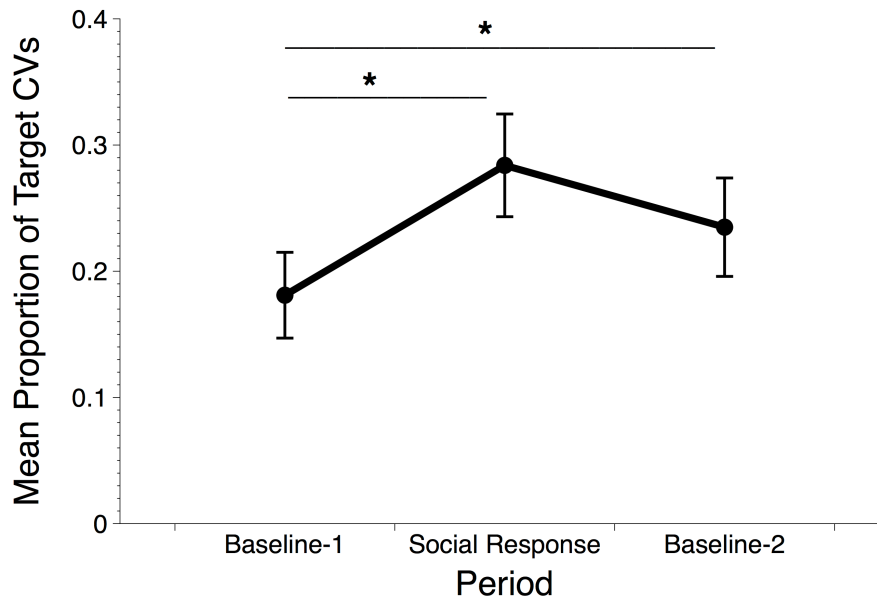


Figure 2.2. Mean proportion of target VCV disyllables (e.g. V[h]V or V[g]V) relative to non-target VCV syllables by period ( $\pm 1$  SE). There was a trend for infants to increase their proportion of target VCV disyllables from baseline to social response,  $p = .075$ .

#### *Proportion of Target CVs to Non-Target CVs*

To further assess changes in infant's target syllable production a 3 (Period: Baseline-1, Social Response, Baseline-2)  $\times$  2 (Condition: [h], [g])  $\times$  2 (Gender) mixed ANOVA was conducted on the proportion of CV syllables that contained the target consonant to total CV syllables produced. There was a significant main effect of Period,  $F(2, 112) = 3.522$ ,  $p < .001$ ,  $\eta_p^2 = .059$  (Figure 2.3). Infants increased their production of target syllables from Baseline-1 to Social Response (Tukey's HSD  $p = .014$ ). There was also a significant main effect of condition,  $F(1, 56) = 14.811$ ,  $p < .001$ ,  $\eta_p^2 = .209$ . Infants exposed to the V[h]V pattern produced a higher

proportion of Target CV syllables than infants in the V[g]V condition. There were no other significant main effects or interactions,  $ps > .131$ .



*Figure 2.3.* Mean proportion of target CV syllables (e.g. [h]v or [g]v) relative to total CV syllables by period ( $\pm 1$  SE). Infants increased their production of target CVs from baseline-1 ( $M = 1.85$ ,  $SD = 2.42$ ) to social response ( $M = 3.53$ ,  $SD = 4.81$ ) and maintained that increase in baseline-2 ( $M = 3.21$ ,  $SD = 4.26$ ).

## Discussion

The results suggest that infants are influenced more by the acoustic regularities than mandibular oscillation patterns when learning novel phonological patterns via socially guided statistical learning. Infants significantly increased their production of both target and non-target VCVs from Baseline-1 to the Social Response period and maintained that increase in Baseline-2 in the absence of caregiver cueing. If infants were primarily constrained by jaw oscillation patterns, we would have expected them to increase only their production of non-Target VCVs. However, infants did learn and produce the novel VCV pattern containing either [h]s or [g]s,

demonstrating that they can learn a new pattern in the absence of mandibular oscillation. Thus, infants appear to learn novel phonological patterns in terms of acoustic patterns rather than motor patterns.

Although infants produced significantly more non-target VCVs in all three periods than target VCVs, they also significantly increased their production of the target pattern. It is not surprising that infants produced more non-target VCVs than target VCVs given there are more possible non-target VCV combinations than there are target VCV combinations. It is possible that production of target VCV syllables were so infrequent that an increase in production from baseline through social response might have been a result of an overall increase vocalizations due to contingent feedback. However, there was a trend for infants to increase their production of target VCV syllables relative to non-target VCV syllables from baseline to social response. Infants also significantly increased their production of target CV syllables (e.g. [h]v or [g]v) relative to the total consonant-vowel utterances produced. Thus, infants learned the target phoneme and increased their production of target CV syllables over and above any increase in CV syllables as a result of contingent feedback.

It is also possible that perhaps infants did not learn the acoustic patterns at all, but rather imitated the phonological patterns in mothers' responses. To demonstrate that infants were learning via SGSL and not imitation we are currently comparing the vowels produced by infants in target VCVs with the vowels contained in the target input of the mothers' immediately preceding response. However, infants continued to produce significantly more target VCV patterned syllables in Baseline-2 as compared to Baseline-1. Thus, in the absence of target input to imitate, infants produced more target VCV patterned utterances. Further, previous experiments on vocal learning have established that infants are not imitating when they learn new

phonological patterns (Goldstein, Syal, & Schwade, under review). When exposed to variable and contingent feedback, infants changed their babbling patterns to match the phonology of the input and did not reproduce the phonemic content of the caregiver input any more than would be expected by chance (Goldstein & Schwade, 2008; Goldstein, Syal, & Schwade, under review). In contrast, infants in the yoked control conditions of these previous studies who did not hear the novel phonological patterns contingent upon their vocalizations did not learn the patterns. Further, if infants learned via imitation of caregiver input, they should have had an easier time imitating caregiver input when the same phonological pattern is repeated. However, infants who received *repetitive* input, either contingently or on a yoked schedule, did not change their babbling patterns (Goldstein, Syal, & Schwade, under review). Taken together, these findings indicate that infants are not learning via imitation but learn the distribution of phonological patterns via social guided statistical learning.

In contrast to the predictions of the Frame/Content theory of speech production, our results demonstrate that at least by 9-months of age, infants are no longer constrained by a dominant mandibular oscillation pattern when producing new sounds. Infants in this experiment specified the phonemic content of their utterances by specifically producing more target V[h]Vs or V[g]Vs in response to contingent caregiver input. These findings have important implications for mechanisms underlying the developmental trajectory of phonological learning. Early infant vocalizations are physiologically constrained until the larynx descends into the throat and the oral cavity expands to increase tongue movement (Kent, 1981). Physiological constraints may explain why 5-month-old infants increase their vocal production when caregivers change their response patterns, but do not modify the quality of their vocalizations (Goldstein, Schwade, & Bornstein, 2009). However, by 7 months of age, infants can modify the content of their

vocalizations in response to contingent input from a social partner (Elkin & Goldstein, 2011). Our results demonstrate that by 9 months of age, motor constraints no longer limit SGS as infants are able to learn phonological patterns that do not require a jaw oscillation. By 9 months of age, infants have several months of practice making CV utterances, so perhaps they have acquired enough experience at this age to learn phonology outside the constraints of the oscillation frame. It is possible that with younger infants an oscillatory frame might govern vocal production. Future studies could examine 7-month-olds' abilities to specify phonemic content as infants this age are just beginning to produce well-formed CV syllables (Oller, 2000) and may be more constrained by jaw oscillation patterns.

While research on statistical learning in vocal production has primarily focused on motor constraints (e.g. MacNeilage, 2008), many other possible constraints outside of the motor domain are yet to be explored. Statistical learning in perception is constrained primarily by limitations on attention and memory (Newport & Aslin, 2004; Toro, Sinnett, & Soto-Faraco, 2005; Conway, Goldstone, & Christiansen, 2007; Saffran 2002; Thiessen, 2011). If statistical learning in production operates via similar mechanisms as in perception, we could predict that limits on attention and memory might also constrain learning in production. Future studies could explore these possible constraints on vocal production by manipulating variables such as the familiarity of the social partner, the signal-to-noise ratio of the social partner's contingent input, or the complexity of the phonological pattern to be learned.

Our results also provide insight into the question of how much variability is necessary for infants to extract the statistical regularities of a pattern. Previous research exposed infants to a VCV pattern in which all three components were varying (i.e. both vowels and the consonant changed) or all three were held constant (Goldstein, Syal, & Schwade, under review). Infants

readily learned the pattern when it was variable, but did not learn from non-varied input. Here, infants learned the novel VCV phonological pattern when the consonant was held constant and only the vowels varied. Thus, infants require some degree of variability to successfully utilize SGSLS but the exact amount of variability required is still unknown. Future studies could vary the amount and position of variability present in contingent feedback to determine the boundary conditions in which infants will and will not detect the pattern.

In summary, our findings demonstrate that infants use socially guided statistical learning to learn novel phonological patterns in terms of the input statistics. When presented with contingent and variable caregiver input, infants are sensitive to low frequency phonemes and rapidly learn to produce the uncommon pattern with only ten minutes of exposure. Finally, by 9 months of age, infants can learn the statistical regularities of phonological input as an acoustic pattern and are not constrained by the mandibular movements necessary to produce the sounds.

## References

- Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16, 1–17.
- Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In B. Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp.353-363). Dordrecht, The Netherlands: Kluwer.
- Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, 297-320.
- Brainard, M. S. & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behavior. *Nature Reviews Neuroscience* 1, 31-40.
- Brainard, M. S. & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, 417, 351-358.
- Conway, C. M., Goldstone, R. L., & Christiansen, M.H. (2007). Spatial constraints on visual statistical learning of multi-element scenes. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society Conference* (pp. 185-190). Austin, TX: Cognitive Science Society.
- Davis, B. L. & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language and Hearing Research*, 38, 1199-1211.
- Doupe, A. J. and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631.
- Eilers, R. E. & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, 124, 199-203.



- Elkin, R. & Goldstein, M. H. (2011). *Titrating out components of the social niche for early vocal learning in prelinguistic infants*. Presented at the 44<sup>th</sup> Annual Meeting of the International Society for Developmental Psychobiology, Washington, D.C.
- Fiser, J. and Aslin, R.N. (2002) Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822–15826.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100, 8030–8035.
- Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *The Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). Oxford: Oxford University Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515–522.
- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants association their own noncry vocalizations with responses from caregivers. *Child Development*, 80(3), 636-644.
- Goldstein, M.H., Syal, S., & Schwade, J.A. (submitted). *Socially guided statistical learning drives the emergence of speech in infants*. Manuscript submitted for publication.
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, 113, 52–58.

- Goldstein, M.H., Waterfall, H.R., Lotem, A., Halpern, J.Y., Schwade, J.A., Onnis, L., & Edelman, S. (2010). General cognitive principles for learning structure in time and place. *Trends in Cognitive Sciences*, 14 (6), 249 - 258.
- Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431-436.
- International Phonetic Association (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Kent, R. D. (1981). Articulatory-acoustic perspectives on speech development. In R. E. Stark (Ed.), *Language behavior and early childhood* (pp. 105-126). New York: Elsevier.
- Kirkham, N Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidences for a domain general learning mechanism. *Cognition*, 83, 835-842.
- Locke, J. L., & Pearson, D. (1990). Linguistic significance of babbling: Evidence from a tracheostomized infant. *Journal of Child Language*, 17, 1-16.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-546.
- MacNeilage, P. F. (2008). *The origin of speech*. Oxford: Oxford University Press.
- MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288, 527-531.
- MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: Frames then content. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp.453-475). Hillsdale, NJ: Lawrence Erlbaum.
- Marler, P. (1970). A comparative approach to vocal learning: song development in White-crowned Sparrows. *Journal of Comparative and Physiological Psychology*, 71, 1-25.

- Mendez, J. M., Dall'Asen, A. G., Cooper, B. G., & Goller, F. (2010). Acquisition of an acoustic template leads to refinement of song motor gestures. *Journal of Neurophysiology*, 104, 984-993.
- Menyhart, O. (2013). An ecological perspective on song learning in the zebra finch (*Taeniopygia Guttata*): The role of social influences. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Nam, H., Goldstein, L. M., Giulivi, S., Levitt, A. G. & Whalen, D. H. (2013). Computational simulation of CV combination preferences in babbling. *Journal of Phonetics*, 41, 63-77.
- Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakerman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences, Early Edition*, 1-6.
- Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of vocal development and disorders. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development* (pp. 509-536). Parkton, MD: New York Press.
- Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology*, 93(2), 260-277.
- Pytte, C. L. & Suthers, R.A. (1999). A bird's own song contributes to conspecific song perception. *NeureReport*, 10, 1773-1778.

- Ross, G. S. (1983). Language functioning and speech development of six children receiving tracheostomy in infancy. *Journal of Communication Disorders*, 15, 95-111.
- Saffran, J. R. (2002). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110-114.
- Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101-105.
- Stoel-Gammon, C. & Herrington, P. B. (1990). Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics & Phonetics*, 4, 145-60.
- Suthers, R. A., Goller, F., & Wild, J. M. (2002). Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proceedings of the National Academy of Sciences*, 99, 5680-5685.
- Thiessen, E. D. (2011). Domain general constraints on learning. *Child Development*, 82, 462-470.
- Toro, J.M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25-B34.
- Vihman, M. M. (1992). Early syllables and the construction of phonology. In *Phonological Development: Models, Research, Implications*, Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon (eds.), 393–422. Timonium, MD: York Press.
- Vihman, M. M. (1993). The construction of a phonological system. In B. Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Eds.), *Developmental*

*neurocognition: Speech and face processing in the first year of life* (pp. 411-419).

Dordrecht, The Netherlands: Kluwer.

Vihman, M. M., & Miller, R. (1988). Words and babble at the threshold of language acquisition. In M. D. Smith, and J. L. Locke (Eds.), *The emergent lexicon: The child's development of a linguistic vocabulary* (pp. 151-183). San Diego: Academic Press.

## CHAPTER 3

### PLAYBACK OF PRELINGUISTIC OBJECT-DIRECTED VOCALIZATIONS FACILITATES WORD LEARNING

The content of this chapter is currently under review at *Developmental Psychobiology*.

### **Abstract**

The present study focuses on a new function of prelinguistic object-directed vocalizations: preparing infants for learning by creating a state of anticipatory readiness. Through rarely studied in human infants, many animals prepare to interact with the environment via self-stimulating behaviors such as vocalizing. We hypothesize that the object-directed vocalizations (ODVs) of human infants function similarly as a form of vocal self-stimulation that promotes learning. Twenty eleven-month-old infants played with a novel object that was labeled immediately after the playback of a recording of the infant's own vocalization. Learning of the word-object association was tested using a preferential looking paradigm. The results suggest that for girls, the playback of an ODV can serve to prepare infants for learning as a form of vocal self-stimulation.

## **Introduction**

Attention is a crucial component of early word learning. Infants show facilitated learning when their attentional focus is coordinated with a social partner who is labeling an object (Baldwin, 1991; Tomasello & Farrar, 1986). The typical social environment of a young infant is filled with rich structure that serves to guide attention in ways that are helpful for learning (Hirsh-Pasek, Golinkoff, & Hollich, 2000; Rickert & Yu, 2010). Caregivers naturally structure their interactions with infants through their speech and actions, providing infants with reliable cues to focus their attention (Goldstein et al., 2010a; Papoušek & Bornstein, 1992). For example, when speaking to infants, caregivers' speech is characterized by exaggerated frequency modulation, slower rate, and hyperarticulation of the vowel space, which serves to increase the saliency of speech as well as attract and maintain infants' attention (Cooper & Aslin, 1994; Fernald & Kuhl, 1987; Kaplan et al., 1995; Singh, Nestor, Parikh, & Yull, 2009; Thiessen, Hill, & Saffran, 2005; Wei, Golinkoff, Houston, & Hirsh-Pasek, 2011; Werker & McLeod, 1989).

Caregivers also tend to move objects in synchrony with their speech (Brand, Baldwin, & Ashburn, 2002). Such motion-speech synchrony attracts infant attention to the object and highlights the association between the label and object (Brand et al., 2002; Brand & Shallcross, 2008; Koterba & Iverson, 2009; Gogate, Bahrick, & Watson, 2000). Parents also engage their infants in joint attention by gesturing and moving objects into the infant's view (Rickert & Yu, 2010), alternating their eye-gaze between the infant and the object (Akhtar & Tomasello, 2000; Baldwin, 1993), and pointing at objects to direct the infant's gaze (Deák, Flom, & Pick, 2000; Flom, Deák, Phill, & Pick, 2004). With experience, infants learn that adults' gaze and pointing behaviors predict the occurrence of interesting objects (Moore & Corkum, 1994; Moore & Povinelli, 2007).



While the efficacy of caregiver cues for organizing infant attention is well established, the infant's own contribution in preparing for word learning in triadic interactions is relatively understudied. Infants may contribute to their own attentional state through the production of immature vocalizations, especially those that are directed at nearby objects. An object-directed vocalization (ODV) is a prelinguistic vocalization produced when an infant is looking at an object that is within reach or being held (Goldstein, Schwade, Briesch, & Syal, 2010). These kinds of vocalizations often receive prompt feedback from caregivers (Goldstein & West, 1999). Such contingent social feedback to babbling is used by infants when building associations between words and objects (Goldstein & Schwade, 2010). For example, when an infant produces an ODV, the caregiver frequently responds to the vocalization by commenting on the object (Albert, Schwade, & Goldstein, 2012), which aligns the caregiver's focus of attention with the infant's. Thus, the infant's own vocalizations serve to structure social interactions in ways that facilitate learning.

Responses to infants' prelinguistic object-directed vocalizations have long-term impacts on children's language development (Goldstein & Schwade, 2010). For example, caregivers often respond to ODVs with *proximal object labels*, which name the object at which the vocalization was directed. Alternatively, caregivers may respond with a label that is phonologically similar to the sound the infant produced. Consider an infant who produces the vocalization "[ba]" while looking at a toy duck. To provide a proximal object label the caregiver could respond by saying "That's a duck!" Conversely, the caregiver could say "Do you want a ball?" thus disregarding the object the infant is attending to and providing a label that is phonologically similar to the sound the infant produced.

Goldstein and Schwade (2010) found a significant positive correlation between the

proportion of proximal object label responses at 9 months and vocabulary size at 15 months, as measured by the MacArthur Communicative Development Inventory (Fenson et al., 1994). In contrast, the proportion of phonologically similar responses at 9 months was *negatively* correlated with vocabulary size at 15 months. These associations between caregiver response types and vocabulary size suggest that not all contingent responses to babbling promote word learning. Infant learning may be slowed when infants hear a label that is incongruent with the object they are looking at. Thus, ODVs may signify a readiness to learn, but word learning is only facilitated when the appropriate object label is provided immediately following the vocalization.

To directly assess the function of ODVs in facilitating associations between words and objects, Goldstein and colleagues tested 11-month-old infants in a word learning paradigm (Goldstein et al., 2010b). One group of infants received a label for a novel object contingently after an ODV and another group received the label after silently looking at the object. Infants were more likely to learn the name for an object if it was labeled after an ODV than after a look alone (Goldstein et al., 2010b).

Why is learning of the word-object association facilitated when infants hear object labels after producing ODVs? Previous work demonstrated that infants' word learning was facilitated when objects were labeled during joint attentional focus with a caregiver (Baldwin, 1991; Tomasello & Farrar, 1986). ODVs may create joint focus by aligning the caregiver's attention with the infant's focus of attention. However, ODVs must also serve an additional function beyond establishing joint attention to explain the results of Goldstein et al. (2010b). In their study, infants in both the ODV and silent look conditions were engaged in joint focus with the experimenter at the time the novel object was labeled. However, only the infants who heard the

label after an ODV learned the word-object pairing. Thus we hypothesize that the act of producing an ODV may be a form of vocal self-stimulation that enhances the infant's receptivity to learning a word-object association. In other words, the acoustic feedback from hearing their own vocalization prepares infants for learning.

Self-stimulation, through rarely examined in infants as a mechanism for organizing attention, has been widely studied in non-human animals. Many species self-stimulate to prepare for exploring, foraging, and mating (Panksepp, 1998). Self-stimulation increases attention to important environmental cues by inducing a state of anticipatory readiness (Panksepp, 1998). For example, vocal self-stimulation is an important component of reproductive behavior in the female ring dove. Females produce nest coos to stimulate ovulation as well as signal to males that mating can ensue (Cheng, 2008; Cohen & Cheng, 1979). However, devocalized females also ovulate in response to playbacks of their own prerecorded nest coo (Cheng, 1992). Thus, simply hearing the vocalization is enough to create a state of anticipatory readiness; producing the sound is not necessary.

In addition to ring doves, vocal self-stimulation is documented in bats to aid in echolocation (Pietsch & Schuller, 1987), rats in anticipation of play with other rats (Knuston, Burgdorf, & Panksepp, 1998), and mallard ducklings to ready a response to the maternal alarm call (Blaich & Miller, 1988). Vocal self-stimulation is also hypothesized by Cheng (1992) to be a mechanism underlying reproductive changes in budgerigars and separation-induced alarm-calls in chickens and puppies. These findings led Locke (1989) and Cheng (1992) to hypothesize that vocalizations produced by human infants may also function as a form of vocal self-stimulation.

In the present study, we investigated the possible function of ODVs as a means of vocal self-stimulation to better understand how social interactions organized by babbling influence

learning. The results of Goldstein and colleagues' (2010b) word-learning study demonstrate that infants' learning of an object-label pairing is facilitated when an object is labeled contingently after an ODV. We hypothesize that when infants produce an ODV, the acoustic feedback of hearing their own vocalizations increases stimulation that prepares infants for learning. Thus, associative learning of objects and their labels is facilitated by a contingent caregiver response to a stimulated infant. However, the results of Goldstein et al. (2010b) left open the possibility that infant ODVs simply signal arousal as the object label was provided after an infant's spontaneous vocalization. By manipulating the timing of when the infant's prerecorded vocalization is played we test the function of ODVs for inducing a state of readiness for learning. If ODVs are a form of vocal self-stimulation, then learning a word-object association should be facilitated when infants hear their own vocalizations played prior to object labeling. Specifically, infants who hear an object label following the playback of their own vocalization should outperform infants who heard the label following a silent look and perform equivalently to infants who heard the label after spontaneously producing an ODV.

## **Method**

### *Participants*

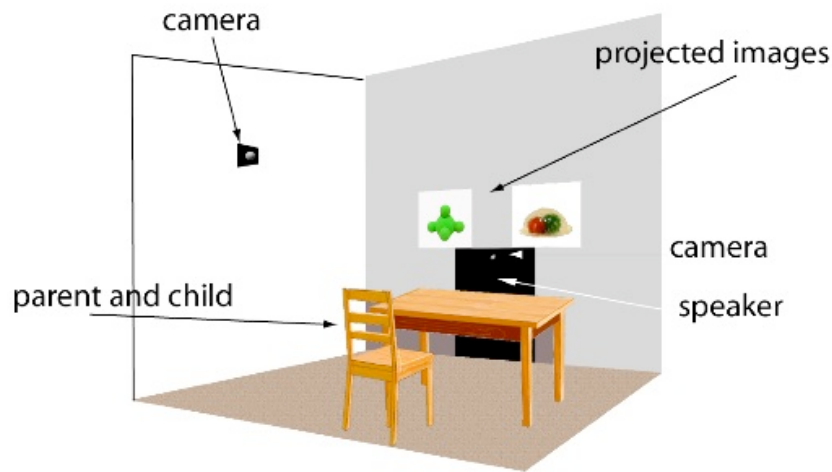
A total of twenty-four 11-month-old infants were tested ( $M = 11;17$ , range 11;6 - 12;18). Data analyses were conducted on a final sample of  $N = 20$  infants (10 boys). Infants in the final sample met an inclusion criterion of silently looking at the object at the moment their vocalization was played on at least five of the six trials during training. Four infants were excluded because, on more than one trial, their gaze was not focused on the object when the prerecorded vocalization played. Infants were primarily from Caucasian, middle-class families. English was the primary language spoken at home. To recruit participants, letters and email

advertisements were sent to families listed in the local newspaper's birth announcements. For their participation, caregivers received an infant t-shirt or bib.

### *Apparatus*

The familiarization period took place in a 3.7-m x 5.5-m playroom containing infant toys. The training and test periods took place in a 3.7-m x 2.4-m room containing two chairs on opposite sides of a .75-m x 1.5-m table. To obtain accurate and detailed recordings of the infants' vocalizations, the infant wore a wireless microphone (Telex FLM-22; Telex Communications, Inc., Burnsville, MN) and transmitter (Telex USR-100) concealed inside the lining of adjustable denim overalls. Infants' behavior was video-recorded from two remote-controlled cameras mounted on the walls of the testing room. The experimenter wore a wireless microphone to record her speech.

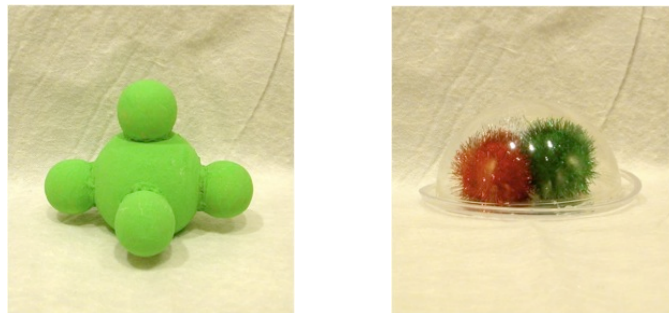
During the preferential looking task, digital photographs of the trained novel objects were presented with Microsoft PowerPoint (2004) on an Apple Powerbook G5 connected to a ceiling-mounted projector (Epson Powerlite 81p; Seiko Epson Corporation, Nagano, Japan; Figure 3.1). The computer was located in a separate room. A video camera positioned approximately 64 in (162.6 cm) from the infant recorded infants' eye gaze. The camera was centered between and below the projected images, and was hidden behind a black curtain. Underneath the camera, a high-quality speaker (RL35; Bang & Olufsen America Inc., Deerfield, Illinois), hidden behind the black curtain, played the infant's prerecorded ODVs during training and played the object labels during test.



*Figure 3.1.* Preferential looking paradigm (similar to that used in Goldstein et al., 2010b). The camera was centered between and below the two projected images, and was hidden behind a black curtain. Underneath the camera, a high-quality speaker hidden behind the black curtain played the infant's prerecorded ODVs during training and played the object labels during test.

### *Stimuli*

*Novel objects and words.* Stimuli were two novel objects differing in shape, texture and color (Figure 3.2). Each age-appropriate object was approximately 15 cm by 20 cm. Results from a previous study (Goldstein et al., 2010b) indicate that 11- and 12-month old infants equally prefer the two objects.



*Figure 3.2.* Novel objects. One object was labeled with a novel word (*riffy* or *koobie*), while the other object was discussed with matched speech but not labeled.

The target object was labeled with one of two novel words: *riffy* or *koobie*. The words

were constructed to match phonological regularities of English nouns (Monaghan, Chater & Christiansen, 2005). Novel labels and which object was named were counterbalanced across participants. During the preferential looking task, photos of the objects were displayed on the wall at 22 in x 19 in (55.0 cm x 48.3 cm, each subtending 19° visual angle), with a center-to-center distance between the images at 35 in (88.9 cm).

*Prerecorded ODVs used in training.* The infant speech syllables used in training were individually recorded in the familiarization period prior to training using Peak Pro 6 software (BIAS Inc., Petaluma, CA). Each infant's most mature object-directed vocalization produced in the familiarization period was used to play back to the infant prior to labeling the novel objects during the training period.

### *Procedure*

Parents brought their infants into the laboratory for a single session divided into three periods: familiarization, training, and test. During the familiarization period, infants and their caregivers engaged in a 15-m free play session with toys in a playroom. Parents were asked to play as they would at home. During the familiarization period, the experimenter noted the quality of each object-directed vocalization the infant produced. Object-directed vocalizations were defined as vocalizations produced while the infant's gaze was fixated on an object within the infant's reach.

To assess vocal maturity, infants' ODVs were classified online according to an infraphonological coding system in which four infraphonological categories are used to describe early vocal development (Oller & Lynch, 1992; Oller, 2000). Infants' most immature vocalizations are *quasi-resonant* vowels, which are vowels produced with a closed vocal tract such as nasal vocalizations and grunts. *Fully-resonant* vowels are vowels that are produced with

an open vocal tract. Infants also produce *marginal syllables*, which are slow sequences of consonant-vowel articulation, with long transitions between consonant and vowel. At 11 months of age, the most mature vocalizations infants tend to produce are fully-resonant vowels combined with faster consonant-vowel transitions (*canonical syllables*, e.g., [ba], [da]). Infants' most mature syllables were selected to use in the training period.

For the training and test periods, infants were seated on their caregivers' laps at a table across from the experimenter in the testing room. Parents listened to music over sound-attenuating headphones and were instructed not to talk or touch the novel toys. The experimenter presented the infant with two equally stimulating novel objects, one at a time. Infants played with each novel object for 30 s prior to hearing their own vocalization (recorded during the familiarization period) played back over speakers at 60dB. The prerecorded vocalizations were played while the infant's gaze was fixated on the object. After the prerecorded vocalization was played, the experimenter immediately labeled the object using infant-directed speech.

The target object was labeled with a novel word embedded in three phrases (e.g. "Look at the *riffy*! See the *riffy*? That's a *riffy*!"). The distracter object was discussed with matched speech that did not contain an object label ("Oh, look at that! See that? Look at that!"). The procedure for the distracter trials was otherwise identical to that of the target trials. While the object was labeled or discussed, the experimenter leaned in and touched the object. The experimenter maintained her gaze on the object except for brief glances at the infant to confirm continued infant attention to the object. If the infant looked away from the object during labeling, the experimenter tapped the toy or said the infant's name to reestablish gaze before continuing labeling. There were six training trials: three trials with one object followed by three trials with



the other object. Thus, infants heard nine phrases with each object. Order of presentation of the target and distracter objects was counterbalanced. The experimenter responded to infants' exploration of the objects with spontaneous nonspecific verbal encouragement (e.g. "Yeah. What do you think?").

Immediately following training, the experimenter left the room, and infants were then tested for their recognition of the object-label pairing using a preferential looking task. During the test period, the caregiver continued to listen to music over headphones and also wore a hat with a black opaque veil. The veil allowed caregivers to see their infants, but prevented their forward view of the preferential looking task.

The preferential looking task began with an animation centered on the screen. When the infant looked at the animation, the task began. The preferential looking task consisted of five 6-s trials, separated by a 2-s inter-stimulus interval of a spinning blue disc with an attention-getting sound to bring the infant's gaze back to midline. The first three trials displayed paired images of familiar objects (e.g. ball, book). The fourth and fifth trials displayed paired pictures of the target and distracter objects. The left/right positions of the pictures were counterbalanced over the two test trials.

In each trial, the pictures were presented silently for 2 s before the onset of the spoken label, recorded in an unfamiliar woman's voice in infant-directed speech. The object label was embedded in the phrase, "Look at the *riffy/koobie*! Can you find it?" Previous tests of infant comprehension during preferential looking tasks have shown that infants more easily recognize words embedded in sentence frames than isolated words (Fernald & Hurtado, 2006). The phrases for the labels *riffy* and *koobie* were approximately the same duration (3,400 ms). The target label began 533 ms after the speech onset, and "Can you find it?" began 2,200 ms after the

speech onset. Each phrase was separated by 1,000 ms of silence. After the offset of speech, there was approximately 1 s of silence before the end of the trial.

### *Coding*

#### Infant Behavior during Training

Infant behavior during the training period was coded for number of object-directed vocalizations, amount of time spent looking at and handling the object, and duration of looking at the experimenter and caregiver. There were no significant differences in infants' looking, holding, or vocalizing behaviors between the target and distracter trials,  $ps > .260$ .

#### Experimenter Behavior During Training

To test for possible differences between the experimenter's behavior during target and distracter trials during training, we observed the experimenter's amount of engagement with the toys, operationalized as the mean number of times the experimenter touched the objects in each trial. The experimenter's interactions with the objects did not differ significantly between the target ( $M = 2.65$ ,  $SD = 1.75$ ) and distracter trials ( $M = 3.18$ ,  $SD = 1.21$ ),  $p = .073$ . Further, the mean number of spontaneous verbal statements made to each object during training did not significantly differ,  $p = .945$ . We also compared the acoustic frequencies of the experimenter's voice during target and distracter object labeling. One of the nine target sentences was randomly selected from training and matched with the same labeling sentence from the distracter object. The minimum, maximum, and mean fundamental frequencies were calculated using speech analysis software (Praat 4.4) and did not differ significantly ( $ts < 1.58$ ,  $ps > .129$ ).

#### Looking at the Objects at Test

Looks to the target and distracter objects during the preferential looking task were coded with frame accuracy using SuperCoder software (Hollich, 2005). All coders were blinded to the location of the target and distracter objects. The first author coded the amount of looking during the preferential looking task to each object. To assess reliability, 33% of sessions were independently re-scored by another coder. Intercoder reliability was  $r = .98$  (range = .95-.98).

To obtain a baseline measure of infants' attention to the two objects during the preferential looking task, the duration of each infant's looking at the target before the onset of the speech stream (the first 2,000 ms of the trial) was divided by the sum of their looking at the target and distracter (Swingley & Aslin, 2007). To measure word comprehension, we calculated looking during a 1430-ms accuracy window that began 367 ms after the onset of the target word (Fernald & Hurtado, 2006; Swingley & Aslin, 2000). The 367-ms interval allows enough time for the infant to initiate an eye movement after the onset of the target word (Fernald & Hurtado, 2006). Infants' looking accuracy was defined as the sum of looking to the target object divided by the sum of their looking to the target and distracter during the accuracy window. For each trial, looking accuracy during the baseline period was compared with looking during the comprehension period and with chance levels (50%).

## **Results**

### *Looking Accuracy*

A 2 (Trial: 1, 2) x 2 (Period: Baseline, Comprehension) x 2 (Gender) mixed ANOVA on looking accuracy found no significant main effects and no significant interaction ( $ps > .082$ ). However, past research suggests that young children may display more variable looking over repeated trials (Schaffer & Plunkett, 1998), perhaps because repeating the same question on

consecutive trials implies an incorrect answer on the first trial and influences looking on later trials (Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). Thus, planned comparisons were used to assess learning separately on Trial 1 and Trial 2 (cf Ballem & Plunkett, 2005).

### *Trial 1*

On trial 1, a 2 (Period: Baseline, Comprehension) x 2 (Gender) mixed ANOVA on looking accuracy showed a significant main effect of period,  $F(1, 18) = 8.835, p = .008, \eta_p^2 = .329$  (Figure 3.3). Infants showed a significant increase in looking to the target object from the baseline period to the comprehension period. This main effect was qualified by a significant Period x Gender interaction,  $F(1, 18) = 8.854, p = .008, \eta_p^2 = .330$ . The interaction was decomposed by Period. Tests of simple main effects revealed that girls were responsible for the increase in looking from the baseline period to the comprehension period,  $F(1, 9) = 28.89, p < .001, \eta_p^2 = .762$ . A Wilcoxon Signed-Ranks test revealed that a significant number of girls ( $n = 9$ ) increased looking from baseline to comprehension,  $T(+) = 54.00, p = .007$ . Boys did not increase their looking to the target from baseline to comprehension,  $p = .785$ . The Period x Gender interaction was also decomposed by Gender to compare performance during the comprehension period. Tests of simple main effects revealed that girls looked more at the target during the comprehension period than did boys,  $F(1, 18) = 5.224, p = .035, \eta_p^2 = .225$ . During the comprehension period, girls' looking to the target was significantly greater than chance performance (50%),  $t(9) = 3.782, p = .004$ . However, girls' performance at baseline was not significantly different from chance,  $p = .144$ . Boys' looking at the target did not exceed chance performance (50%) during either period,  $ps > .6$ . There was no significant effect of gender at baseline,  $p = .131$ .

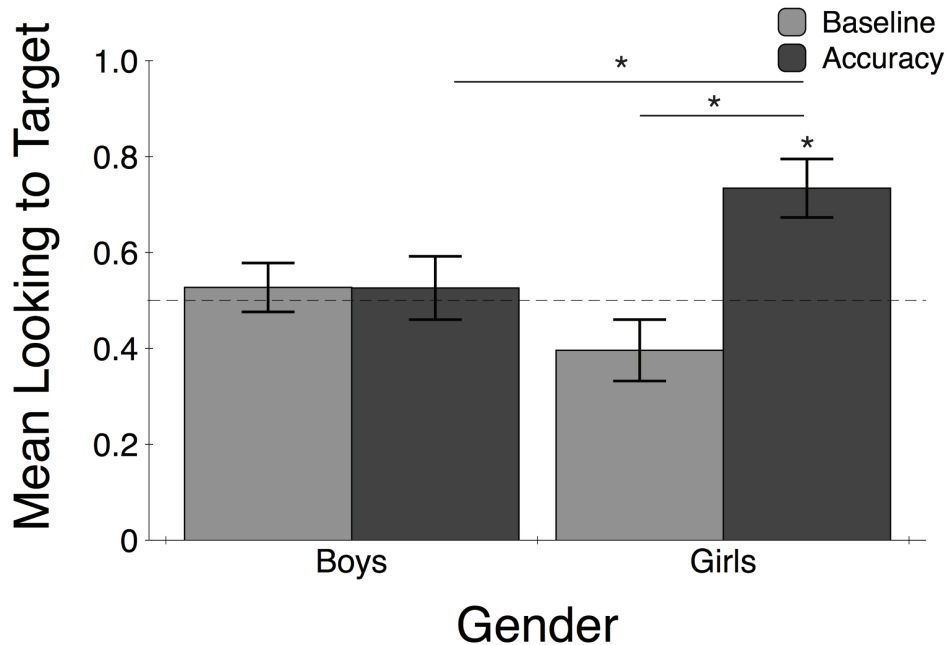


Figure 3.3. Mean proportion of looking to the target on Trial 1 by comprehension period and gender ( $\pm 1$  SE). On trial 1, neither girls ( $M = .396$ ,  $SD = .20$ ) nor boys ( $M = .527$ ,  $SD = .16$ ) showed a preference for the target object at baseline and looking at baseline was at chance (chance performance = .5, as indicated by the dashed line). During the comprehension period, girls looked longer at the target object ( $M = .698$ ,  $SD = .18$ ) than did boys ( $M = .498$ ,  $SD = .19$ ). Girls showed a significant increase in looking to the target object from baseline to comprehension period and girls' performance was significantly better than chance performance.

### Trial 2

On Trial 2, a 2 (Period: Baseline, Comprehension) x 2 (Gender) mixed ANOVA on looking accuracy found no significant main effects and no significant interaction ( $ps > .79$ ).

### Comparison of learning to previously published controls

Previous work by Goldstein and colleagues (2010b) found that infants who received the object label after a silent look did not show a reliable preference for the named object during comprehension,  $p = .46$ . Only 8 of 20 infants in the silent control condition increased looking

from baseline to comprehension. However, a significant number of infants ( $n = 13$  of 20) who heard the object labeled promptly after babbling demonstrated learning by increasing their looking from baseline to comprehension,  $T(+) = 125.00, p = .003$  (from Goldstein et al., 2010b). In the current experiment, 12 of 20 infants who heard their own prerecorded vocalization showed a trend toward increased looking from baseline to comprehension  $T(+) = 156.00, p = .056$ . Nine of 10 girls demonstrated learning by increasing their looking from baseline to comprehension,  $T(+) = 54.00, p = .007$ . Thus, at least for girls, hearing a recorded version of their own vocalization facilitates learning over and above any effect of receiving the label after a silent look (as evidenced by no learning by the silent control condition in Goldstein et al., 2010b). Further, girls who heard their own prerecorded vocalization showed a reliable preference for the target object, as did infants in the ODV condition who produced the vocalization themselves (from Goldstein et al., 2010b).

#### *Comparison of training behaviors by gender*

Given the gender differences observed in looking accuracy, independent samples t-tests were conducted to assess gender differences in infant behaviors during training (all Bonferroni-corrected  $\alpha s = .025$ ). There was no significant gender difference in the number of trials during which the experimenter reestablished the infant's gaze on the object during labeling,  $p = .51$ . Also, the number of trials in which the experimenter reestablished infant gaze was not significantly correlated with infant looking to the target object during the comprehension period at test,  $p = .383$ . Further, there were no significant gender differences in time spent looking at or holding the objects on either the target or distracter trials,  $p s > .4$ . There were also no significant gender differences in time spent looking at the experimenter on either the target or distracter trials,  $p s > .054$ . There were no significant gender differences in the number of vocalizations

directed at the objects on either the target or distracter trials,  $ps > .073$ . Finally, the experimenter's behavior during training did not differ between genders, either in amount of speech ( $ps > .870$ ), mean, maximum and minimum pitch during labeling ( $ps > .06$ ) or interactions with the training objects ( $ps > .164$ ).

### *Quality of Playback Vocalizations*

To compare the quality of the infant vocalizations played back during training, infants' vocalizations were classified into one of four infraphonological categories used to describe early vocal development (see Procedure; Oller & Lynch, 1992; Oller, 2000). The distribution across categories was roughly equal between boys and girls. Six girls and seven boys heard a canonical syllable, while two girls heard marginal syllables. Two girls and three boys heard fully-resonant vowels. Thus, the quality of the sounds heard in training prior to the object being labeled was approximately equivalent between genders.

## **Discussion**

For girls, labeling an object immediately after playback of the infant's own ODV facilitated learning of the word-object pairing. Girls in the current study show a reliable preference for the target object, whereas infants who heard the label after a silent look to the object did not demonstrate learning (Goldstein et al., 2010b). Further, a similar number of infants learned in the two conditions (12 of 20 in the current experiment versus 13 of 20 in the ODV condition of Goldstein et al., 2010b). Specifically, girls in the current study show comparable learning to infants who heard the object labeled after producing an ODV (Goldstein et al., 2010b). These results lend support to our hypothesis that an infant's own vocalization can function as a form of vocal self-stimulation to increase their readiness to learn. In our view, an

infant's object-directed vocalization creates a state of receptivity for learning at the moment the caregiver is likely to label the object the infant is attending to. Thus, learning of the word-object association is facilitated when caregivers provide a label contingently upon an ODV. Although vocal self-stimulation has been documented in several species (Blaich & Miller, 1988; Cheng, 2008; Knuston et al., 1998), the present study is the first to demonstrate a function of vocal self-stimulation in human infants.

We were surprised that only the girls in this experiment showed reliable learning of the word-object pairing. The gender difference is not readily explained by differences in behavior during training, as girls and boys did not differ in the amount of time spent looking at the training objects or the experimenter. Infants also did not differ in the amount of time spent holding the objects. Also, while labeling the object in training, girls and boys did not differ in the number of trials the experimenter refocused the infant's gaze on the object. Further, the quality of the vocalization played back to the infants during training did not differ between genders. Finally, the experimenter's behavior in terms of her amount of speech, acoustic frequency, and interactions with the objects did not differ between genders. However, past research with similarly-aged infants demonstrates that girls sometimes outperform boys on word learning tasks (Woodward, Markman, & Fitzsimmons, 1994).

Our findings offer an explanation for why learning improves when infants hear object labels after producing ODVs. If an infant's vocalization induces a state of anticipatory readiness, it follows that learning of the information presented immediately following the vocalization should be enhanced. Our interpretation complements the results of Goldstein and colleagues (2010b) who demonstrated that learning word-object associations is facilitated when an object is labeled after an ODV, but extends the study by demonstrating that a state of readiness can be



induced using the playback of an infant vocalization. Infant production of the vocalization is not necessary to create this readiness. However, the anticipatory state created by infants' vocalizations may help or hinder learning depending on the relevance of the information provided by the caregiver. Such context-dependence explains the relations between the appropriateness of labels that parents provided following ODVs and the direction of correlation with later vocabulary size (Goldstein & Schwade, 2010). Therefore, social interactions organized by ODVs are likely to have more salience than interactions at other points in time.

An outstanding question concerns the mechanism underlying the facilitative effects of ODVs. One possibility is that the infant's vocalization is arousing. Two processes could explain the role of arousal in ODVs. Our results suggest that ODVs may function as a catalyst for increasing arousal as a type of vocal self-stimulation (*self-stimulating hypothesis*). Vocalizing may modulate infants' arousal level in response to perceptual stimulation. Alternatively, infants may experience a general increase in arousal when looking at an object, which stimulates vocal production (*arousal-signaling hypothesis*). If this is the case, an ODV may be an indicator of a readiness to learn rather than the catalyst. To test these hypotheses, future studies should incorporate physiological measures to assess the time course between vocalization and arousal. Galvanic skin response (GSR), a technique for measuring changes in skin conductance that result from increasing electrodermal activity of the autonomic nervous system, would be an appropriate measure. Recently, GSR was used to measure arousal with 5-month-old infants after the presentation of an auditory stimulus (i.e. clapping) (Ham & Tronick, 2008).

In a future study, GSR could be used to precisely test the relationship between arousal and the production of ODVs. If the arousal-stimulating hypothesis is supported, then prelinguistic vocalizations are responsible for an increase in arousal. Thus social responses to

ODVs that incorporate object labels should facilitate learning. However, if the arousal-signaling hypothesis is supported, ODVs may be a byproduct of an increase in arousal. Therefore, arousing an infant with an external stimulus prior to labeling (e.g., an exciting sound) should also facilitate learning. Regardless of which hypothesis is supported, understanding the time course of arousal will enhance our understanding of the function of ODVs as a mechanism for learning.

An additional question in relation to the arousal-signaling hypothesis regards the degree of specificity of the vocalization necessary to facilitate learning. Can the effects of ODVs be replicated using other types of sounds or is the effect specific to the infant's own vocalizations? Results from the non-human animal literature suggest that the effects of self-stimulation can be elicited using prerecorded vocalizations of conspecifics in birds and rats (Cheng, 1992; Knuston et al., 1998). However, the breadth of the facilitative effects of vocal self-stimulation in humans for learning word-object associations is yet untested. If vocal self-stimulation in human infants functions in a similar manner as in other species, then playback of another infant's vocalization should also create a state of readiness.

A follow-up study is currently in progress to assess which acoustic characteristics of vocalizations may induce a state of anticipatory readiness. Can the effects of ODVs be replicated using other types of sounds or is the effect specific to the infant's own vocalizations? One hundred 11-month-old infants participated in one of five conditions (well-formed syllable, infant babble, canonical syllable, non-speech sound, silent control). The well-formed syllable was a prerecorded canonical syllable (e.g. [ba]) spoken by a female adult. The infant babble was the infant's own prerecorded vocalization, collected prior to training. The canonical syllable was a canonical syllable produced by an infant in the infant babble condition. The non-speech sound was recorded from a squeaky toy. Infants were presented with two novel objects (target,

distracter). When the infant looked at the object, the condition-specific sound was played (no sound was played in the silent control condition). The object was then labeled using three phrases [e.g. "Look at the riffy!"(target); Look at that."(distracter)]. Learning was measured by infants' preferential looking to the target when hearing the label. Preliminary results suggest that only infants in the well-formed syllable condition showed reliable learning. However, in the infant babble condition, learning was positively correlated with the speech-like quality of the babble. Three adults rated each vocalization on a seven-point scale of how speech-like it sounded and mean speech ratings were calculated for each vocalization. Infants who heard their own speech-like babble looked more to the target object than infants who heard their own immature vocalization.

These additional results suggest that by 11 months, infants have learned that speech-like sounds have signal value in predicting contingent adult interaction. Previous work suggests that word recognition is facilitated when objects are labeled in a full sentence rather than in isolation (Fernald & Hurtado, 2006). However, these results are the first to suggest that a single isolated syllable heard prior to labeling also facilitates learning. According to the social-gating hypothesis, social interaction organizes infant attention and increases arousal, thereby facilitating learning (Kuhl, 2007). The results support the concept of social gating by suggesting that the infant's own speech-like sounds create a readiness to learn.

Finally, our findings of infant contributions to learning via ODVs call for a more integrative approach to understanding the parent-infant interactions that support language development. Caregiver cues, such as infant-directed speech, motionese, and pointing, are clearly important in organizing infant attention for learning, but how are these cues influenced by ODVs? This new function of prelinguistic vocalizing should be integrated into new studies that

measure parental responsiveness to ODVs and infant learning *at the same time* to capture the richness of early socially-embedded learning.

## References

- Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, and G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 114–135). New York: Oxford University Press.
- Albert, R. R., Schwade, J. A., & Goldstein, M. H. (June, 2012). *Mothers versus others: How caregiving experience influences caregiver-infant interactions*. Poster presented at the biennial meeting of the International Society of Infant Studies, Minneapolis, MN.
- Baldwin, D. A. (1991). Infants' contribution to achievement of joint reference. *Child Development*, 62, 874-890.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32, 159–173.
- Blaich, C. F. & Miller, D. B. (1988). Alarm call responsivity of mallard ducklings: VI. Effects of sibling and self-produced auditory stimulation. *Journal of Comparative Psychology*, 102, 56-60.
- Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, 5, 72-83.
- Brand, R. J. & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, 11, 853-861.
- Cheng, M. -F. (1992). For whom does the female dove coo? A case for the role of vocal self-stimulation. *Animal Behaviour*, 43, 1035–1044.
- Cheng, M. -F. (2008). The role of vocal self-stimulation in female responses to males: Implications for state-reading. *Hormones and Behavior*, 53, 1-10.

- Cohen, J. & Cheng, M.-F. (1979). Role of vocalization in the reproductive cycle of ring doves (*Streptopelia risoria*): effects of hypoglossal nerve section on the reproductive behavior and physiology of the female. *Hormones and Behavior*, 13, 113-127.
- Cooper, R. P. & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, 65, 1663-1677.
- Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36, 511-523.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59.
- Fernald, A. & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 279-293.
- Fernald, A. & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9, F33-F40.
- Flom, R., Deák, G. O., Phill, C., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, 27, 181-194.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100, 8030-8035.
- Goldstein, M. H., & Schwade, J. A. (2008). Social interaction to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515-523.

- Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *The Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). Oxford: Oxford University Press.
- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants association their own noncry vocalizations with responses from caregivers. *Child Development, 80*, 636-644.
- Goldstein, M. H., Schwade, J. A., Briesch, J., & Syal, S. (2010b). Learning while babbling: prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy, 15*, 362-391.
- Goldstein, M.H., Waterfall, H.R., Lotem, A., Halpern, J.Y., Schwade, J.A., Onnis, L., & Edelman, S. (2010a). General cognitive principles for learning structure in time and place. *Trends in Cognitive Sciences, 14*, 249 - 258.
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology, 113*, 52–58.
- Gogate, L. J., Bahrick, L., E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71*, 878-894.
- Ham, J. & Tronick, E. (2008). A procedure for the measurement of infant skin conductance and its initial validation using clap induced startle. *Developmental Psychobiology, 50*, 626-631.

Hirsh-Pasek, K., Golinkoff, R. M., & Hollich, G. (2000). An emergentist coalition model for word learning: mapping words to objects is a product of the interaction of multiple cues. In R. Golinkoff, K. Hirsh-Pasek, L. Bloom, G. Hollich, L. Smith, A. L. Woodward, Akhtar, L., Tomasello, M., and Hollich, G. (Eds.) *Becoming a word learner: A debate on lexical acquisition*, (pp. 136-164). Oxford: Oxford University Press.

Hollich, G. (2005). SuperCoder: A program for coding preferential looking (Version 1.5).[Computer Software]. West Lafayette, IN: Purdue University.

Kaplan, P. S., Goldstein, M. H., Huckleby, E. R., & Cooper, R. P. (1995). Habituation, sensitization, and infants' responses to motherese speech. *Developmental Psychobiology*, 28, 45-57.

Koterba, E. A. & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development*, 32, 437- 444.

Knuston, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology*, 112, 65-73.

Locke, J. (1989). Babbling and early speech: Continuity and individual differences. *First Language*, 9, 191-205.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, 11, 628-634.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, 96, 143–182.



- Moore, C. & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review, 14*, 349-372.
- Moore, C. & Povinelli, D. J. (2007). Differences in how 12-and 24-month-olds interpret the gaze of adults. *Infancy, 11*, 215-231.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.
- Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of vocal development and disorders. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development* (pp. 509-536). Parkton, MD: New York Press.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY: Oxford University Press.
- Papoušek, H. & Bornstein, M. H. (1992). Didactic parental support for human infants. In H. Papoušek, U. Jürgens, and M. Papoušek (Eds.), *Nonverbal Vocal Communications* (pp. 209-229). New York: Cambridge University Press.
- Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connection Science, 20*, 73-89.
- Pietsch, G. & Schuller, G. (1987). Auditory self-stimulation by vocalization in the CF-FM bat, *Rhinolophus Rouxi*. *Journal of Comparative Physiology A, 160*, 635-644.
- Rickert, M. E. & Yu, C. (2010). *Joint attention through the hands: Investigating the timing of object labeling in dyadic social interaction*. Paper presented at the IEEE 9<sup>th</sup> International Conference on Development and Learning (ICDL), 114-119.
- Schaffer, G. & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*, 309-320.

- Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, 14, 654-666.
- Smith, L. B. (2000). How to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. Smith, A. Woodward, N. Akhtar, M. Tomasello, and G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). New York: Oxford University Press.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99–132.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 53-71.
- Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57, 1454-1463.
- Tomasello, M. & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4, 187-211.
- Weiyi, M., Golinkoff, R. M., Houston, D. & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7, 209-225.
- Werker, J. F. & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, 43, 230-246.
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553-566.

Yu, C. & Ballard, D. H. (2007) A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70, 2149-2165.

## CHAPTER 4

### DETERMINANTS OF CAREGIVER RESPONDING TO INFANT VOCALIZATIONS: FINDINGS FROM THE PLAYBACK PARADIGM

## **Abstract**

Caregivers' contingent responses to prelinguistic vocalizations have both immediate and long-term effects on infants' speech and language development. Despite the importance of early caregiver responses for later learning, the aspects of infant behavior that drive responsiveness have not been systematically examined. In addition, it is unknown how caregiver responsiveness develops. The present studies manipulated auditory and visual features of vocalizing infants to titrate out the characteristics that influence caregiver responsiveness. In addition, we examined the influence of caregiving experience on responding to infant behavior. The current experiments utilized a playback paradigm, a method widely used in studies of animal communication. Female caregivers reacted to prerecorded audiovisual playbacks of unfamiliar infants' vocalizations and actions. Experiment 1 validated the playback paradigm as a measure of caregiver behavior. Mothers interacted with their 9-month-old infants in a play session and then responded to the playback stimuli. In both contexts, mothers responded to a similar proportion of infant vocalizations and in similar ways. Experiment 2 specified the acoustic determinants of caregiver responses to infant vocalizations and assessed the role of caregiving experience on women's responses. The acoustic qualities and the directedness of the vocalizations were manipulated to systematically examine their effects on women's responses. Women were more likely to respond to mature vowel vocalizations than immature vowels. For consonant-vowel vocalizations, women provided different types of responses depending upon the maturity of the syllable. Women also responded more to object-directed vocalizations than undirected vocalizations. Further, caregiving experience influenced women's ratings of the speech-like qualities of infant vocalizations. The results of these experiments validate the playback paradigm for measuring caregiver responses to infant vocalizations and are the first to demonstrate that the acoustic

characteristics and context of infant vocalizations reliably influence social responses from caregivers.

## **Introduction**

Language development is traditionally studied primarily from the infant's perspective, examining caregiver input only as it predicts developmental outcomes. An infant-centered approach has revealed the importance of caregiver responsiveness to infant behavior for early language development (Landry, Smith, Miller-Loncar & Swank, 1997; Bornstein & Tamis-LeMonda, 1989; Baumwell, Tamis-LeMonda, & Bornstein, 1997; Hart & Risley, 1995; Rollins, 2003; Tamis-LeMonda, Bornstein, & Baumwell, 2001). In particular, caregivers' responses to infants' prelinguistic vocalizations are predictive of later receptive and productive vocabulary. Caregivers' descriptions and imitative responses to 9- and 13-month-old infants' vocalizations are positively correlated with several language milestones including infants' production of first words and multi-word utterances (Tamis-LeMonda, Bornstein, & Baumwell, 2001).

One limitation of the infant-centered approach, however, is that the details and development of caregiver behavior is relatively understudied. Exactly what aspects of moment-to-moment interactions between caregivers and infants facilitate language development? Behaviors that organize visual attention during labeling constitute a reliable cue that may promote word learning. For example, parents engage their infants in joint attention by gesturing and moving objects into the infant's view (Rickert & Yu, 2010) and pointing at objects to direct the infant's gaze (Deák, Flom, & Pick, 2000; Flom, Deák, Phill, & Pick, 2004). In fact, the object in or referred to by a caregiver's hands is a better predictor of what the caregiver is talking about than the focus of the caregiver's gaze (Yu et al., 2013).

Caregivers' reactions to infant vocalizations are also a reliable source of social information for language learning. When infants produce prelinguistic vocalizations, caregivers consistently respond (Goldstein & West, 1999; Papoušek, 1989). Mothers provide a wide range of responses

to infant vocalizations such as labeling objects, imitating, describing the infant's behaviors, asking questions, or redirecting the infant's attention (Paavola, Kunnari, Moilanen, & Lehtihalmes, 2005; Tamis-LeMonda Bornstein, & Baumwell, 2001). In turn, infants modify their vocalizations in response to caregivers' contingent feedback (Goldstein, King, & West, 2003). Caregivers' contingent responses to babbling also increase infants' production of speech-like vocalizations and facilitate learning new phonological patterns (Goldstein et al., 2003; Goldstein, Syal & Schwade, submitted). Further, infants show facilitated learning of word-object associations when an object is labeled immediately after the infant babbles at it (Goldstein et al., 2010b). Thus, caregivers' responses to infants' prelinguistic vocalizations have immediate effects on language learning (Goldstein & Schwade, 2010).

Not all forms of caregiver responsiveness have the same impact on infant learning. Caregivers' sensitive responses such as providing object labels (Stevens, Blake, Vitale & MacDonald, 1998) or asking questions (Furrow, Nelson, & Benedict, 1979; Gleitman, Newport, & Gleitman, 1984) have positive relations with later language development. In contrast, redirective responses negatively impact later vocabulary (Akhtar, Dunham & Dunham, 1991; Della Corte, Benedict & Klein, 1983; Tomasello & Farrar, 1986). Further, the type of labels that mothers provide after vocalizations can differentially influence later vocabulary. For example, maternal labeling of objects that 9-month-old infants babble at is positively correlated with vocabulary at 15 months. In contrast, labeling absent objects in response to object-directed vocalizations is negatively correlated with later vocabulary (Goldstein & Schwade, 2010). Infant learning may be slowed when infants hear a label that is incongruent with the object they are looking at. The varied associations between different caregiver response types and vocabulary size suggest that a systematic examination of the forces driving responsiveness is needed.



The present experiments focus on understanding the development of caregiver responding to noncry infant vocalizations. Prior work has shown that early vocal learning is facilitated by social feedback (e.g. Goldstein & Schwade, 2008); here we investigate the ways in which babbling organizes the social interactions that lead to feedback. To achieve a high degree of control over the form and timing of infant behavior that will serve as stimuli, we utilized a playback paradigm, which is a method widely used in studies of animal communication. We recorded and recombined infant vocalizations and actions to titrate out their specific effects on caregiver responses.

Our stimuli were representative of the wide range of vocalizations that infants produce in their first year. Prelinguistic vocalizations can be categorized based on their acoustic properties using an infraphonological coding system (Oller, 2000). The system describes the major changes in vocal production over the first year using four major syllable types: quasi-resonant vowels (QR), fully-resonant vowels (FR), marginal syllables (MS), and canonical syllables (CS; Oller & Lynch, 1992; Oller, 2000). From birth, infants produce QR vocalizations, which are vowels produced with a closed vocal tract. These vocalizations include nasal vocalizations and grunts. At 3–8 months, the vocal tract becomes more open and fully resonant sounds are produced. Infants then produce marginal syllables, which are slow sequences of consonant–vowel articulation, with long transitions ( $> 200$  msec) between the consonant and vowel. From 5 to 10 months, infants begin to produce fully resonant vowel nuclei combined with and faster transitions between consonants and vowels, resulting in canonical syllables (e.g., [ba], [da]) (Oller, 2000). At 9 months of age, most infants regularly produce all four infraphonological types (Oller et al., 1999).

In addition to vocal quality, we manipulated the typical contexts in which vocalizations are produced. In play interactions between mothers and infants, the infants' vocalizations are usually aimed in one of three ways: caregiver-directed, object-directed, or undirected. Caregiver-directed vocalizations (CDVs) are those produced while the infant is looking at the mother's face. At 5 months of age, mothers' responses to infants' caregiver-directed vocalizations are positively correlated with later vocabulary at 18 months (Goldstein, Schwade, & Kirkpatrick, 2013). Infants also direct their vocalizations at objects. Object-directed vocalizations (ODVs) are defined as those produced while the infant is looking at an object that is held or within reach. Parental responsiveness to their 9-month-old infants' ODVs predicted infants' language development at 15 months (Goldstein & Schwade, 2010). Furthermore, 11-month-old infants are more likely to learn the name for an object if it is labeled after an ODV rather than after a look alone (Goldstein, Schwade, Briesch, & Syal, 2010). Finally, undirected vocalizations (UDVs) are vocalizations produced at neither an object nor a caregiver and can be characterized as vocalizing into empty space. Caregivers may respond differently to infant vocalizations based on their directedness. For example, ODVs may elicit more object labels and specific interactions related to the focus of the infant's attention because the vocalization attracts the caregiver's attention to the place the infant is already attending. In contrast, UDVs may encourage caregivers to engage in more conversational turn-taking by responding with descriptions of the infant's behavior or placeholders (e.g. "Oh yeah" or "Uh-huh") that acknowledge the infant's contribution to the protoconversation without providing specific information.

Vocal quality and context of infant vocalizations may interact to drive patterns of adult responsiveness, thus we systematically manipulated infant vocal quality and directedness using the playback paradigm. An additional strength of the paradigm is that each participant responded

to the same instances of infant vocalizations. In live play sessions, it is difficult to isolate the specific qualities of infant vocalizations that influence caregivers' responses because infants frequently produce many vocalizations in quick succession. However, using the playback paradigm, participants responded to a controlled set of stimuli consisting of single syllables, which allowed us to isolate characteristics of infant vocalizations that impact caregiver responses. The stimulus set also consisted only of unfamiliar infants, which eliminated potential response biases due to caregivers' idiosyncratic histories with their own infants.

Playback studies are widely used in the animal communication literature to assess the potency of vocal communication for the receiver. For example, playbacks have been used to assess baboons' responses to alarm calls (Kitchen et al., 2010), maternal responses to piglet screams (Illman et al., 2008), Japanese macaques' maternal responses to infant calls (Shizawa et al., 2005), and female cowbirds' preferences for males' song (Smith, King, & West, 2000). Playbacks have also been used to assess human caregivers' responses to infant behavior. Most playback studies using human infant vocalizations have been restricted to examining adult responses to fusses and cries (Green, Jones, Gustafson, 1987; Gustafson & Green, 1989; Wood & Gustafson, 2001; Zeskind, Klein, & Marshall, 1992; Zeskind & Marshall, 1988). The few experiments that examined caregivers' responses to babbling had adults rate characteristics of infants, such as happiness or attractiveness, rather than specify their own behavioral responses to the vocalizations (Bloom & Lo, 1990; Bloom, D'Odorico, & Beaumont, 1993; Papoušek, 1989). The only published study that examined caregivers' reactions to prelinguistic vocalizations using a playback paradigm asked caregivers to interpret the infant's internal state (e.g., "I think the infant wants a toy"), rather than provide their own response to the vocalization (Goldstein & West, 1999).

The present studies had caregivers provide in-the-moment responses indicating their reaction to prerecorded examples of infant behavior. Caregivers imagined they were in the room interacting with the infants shown on video and provided two types of responses to each example of infant behavior. Caregivers vocally responded to the infant's behavior as if they were actually interacting with the infant. Caregivers also responded to the infants via a computer-driven digitizing tablet and stylus, positioned just below the monitor displaying the examples of infant behavior. The tablet contained a distance scale with 'Stay where you are' and 'Go to baby' as endpoints. Caregivers indicated their desired proximity to the infant by touching the tablet with a stylus.

The first experiment served to validate the playback paradigm as a measure of caregiver behavior by comparing caregivers' reactions to the playback stimuli to those given to their own infants during play. Past research demonstrates that mothers show stable patterns of responding across contexts (e.g., in a laboratory environment versus at home; Crockenberg & Litman, 1990; Rothbaum & Crockenberg, 1995). Mothers also show consistent response patterns to babbling whether playing with their own familiar infant or an unfamiliar infant in a live play session (Albert, Schwade, Goldstein, 2013). Additionally, mothers respond naturally, using infant-directed speech, when interacting with their infants over video (Smith & Trainor, 2008). These findings indicated that mothers should respond naturally and consistently to audio/video stimuli of unfamiliar infants.

In Experiment 2, we used the playback paradigm to specify the acoustic and contextual determinants of caregivers' responses to prelinguistic vocalizations. We compared caregivers' responses to changes in resonance (quasi- vs. fully-resonant vowel sounds) and changes in the timing of consonant-vowel (CV) transitions (marginal vs. canonical syllables) to determine how

these acoustic features drive caregivers' responses. Further, we compared caregivers' responses to changes in the context in which the vocalization was produced (object-directed vs. undirected) to determine how the directedness of infant vocalizations impacts caregiver responding. Previous work by Gros-Louis and colleagues (2006) suggested that caregivers respond more often to more speech-like vocalizations when in face-to-face interactions with their infants. Thus we predicted that more mature vocal types (FRV, CS) should receive more vocal reactions that involve attending to the infant and stronger proximity responses (e.g. indicated that they would move closer to the infant). We also predicted that caregivers should respond more to vocalizations in an object-directed context than in an undirected context because vocalizations directed at objects give caregivers a specific opportunity to provide immediately relevant information to infants.

After collecting caregivers' immediate reactions to infant behavior, we asked them to rate each vocalization on a 7-point Likert scale for how speech-like it sounded. We then compared their ratings to assess whether the infraphonological categories predict caregivers' perceptions of well-formed syllables. We predicted that caregivers would respond more often to syllables with mature acoustic characteristics. Previous playback studies found that mothers rated fully-voiced sounds as more enjoyable and as communicating more intent than quasi-resonant vocalizations (Beaumont & Bloom, 1993; Bloom et al., 1993). Also, mothers interpreted more speech-like sounds as indicators that an infant wants something (Goldstein & West, 1999). Therefore, FRV vocalizations should be perceived as more speech-like and receive stronger proximity responses than QRV vocalizations. Caregivers may also rate canonical syllables as more speech-like than marginal syllables, as previous work demonstrates that caregivers readily recognize canonical syllables when exposed to samples of infant vocalizations (Oller, Basinger, & Eilers, 1996). If

caregivers perceive differences between the two forms of consonant-vowel stimuli, they should rate CS stimuli as more speech-like than MS stimuli.

We also assessed the role of caregiving experience on responsiveness. Most studies using adult ratings of infant sounds have failed to find an effect of caregiving experience (Bloom et al., 1993; Papoušek, 1989). However, these studies did not assess responsiveness; instead they asked adults to either rate infant sounds on scales indexing emotional content (Papoušek, 1989) or rate the attractiveness of the infants (Bloom et al., 1993). In contrast, studies of caregivers' responses to cries and fusses did find experience-related differences (Giardino et al., 2008; Green, Jones, & Gustafson, 1987). The present study is the first to assess the role of caregiver experience in responding to (rather than rating) non-cry prelinguistic vocalizations. In studies of live interaction, women with varying amounts of caregiving experience differed in their interactions with infants. Mothers were more selective when responding to infants' vocalizations and were more sensitive to the infant's focus of attention than non-mothers (Albert, Schwade, & Goldstein, 2012).

In the present experiments, we used participants' parity as a measure of caregiving experience. Parity refers to the number of times a woman has given birth. We compared the responses nulliparous women (non-mothers) with the responses of primiparous women (first-time mothers) and multiparous women (mothers with two or more children). We predicted that experienced participants should react to mature syllables (FR, CS) more strongly than to immature syllables (QR, MS). Mature syllable types should also receive closer proximity scores and higher proportions of vocal response types that involve attending to the infant. The speech-like ratings of the stimuli from experienced caregivers should also be more sensitive to syllable type.

## **Experiment 1**

### **Method**

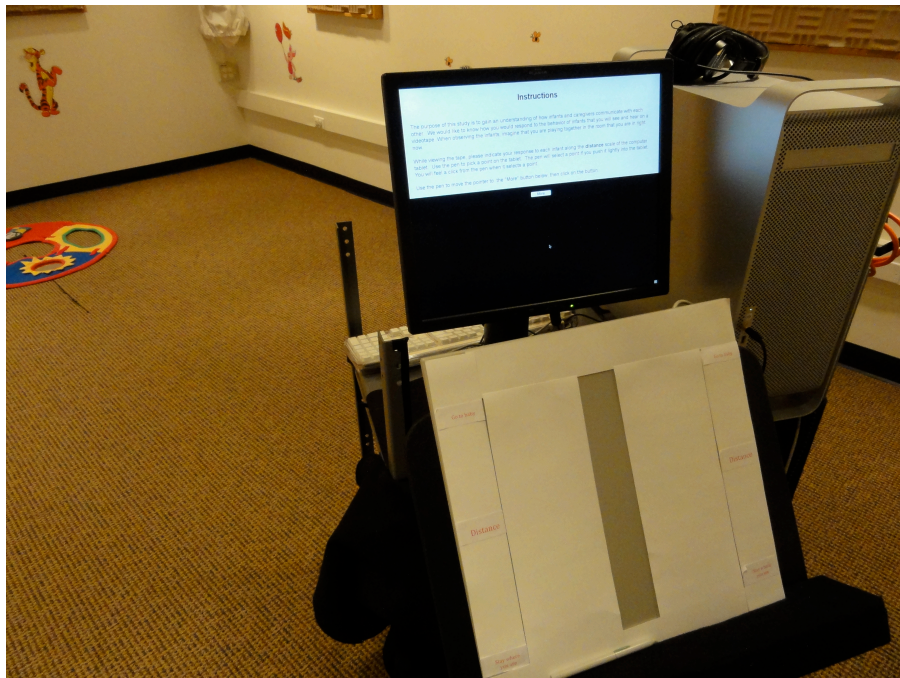
#### *Participants*

Forty mothers (n= 20 primiparous) participated with their 9-month-old infants (20 male; mean age 9 months 24 days; range 8 months 21 days-10 months 15 days). An additional 16 dyads were tested but excluded because the infant cried or fussed excessively (n= 7), the caregiver failed to follow directions (n= 4) or speak in English (n= 2), or there was an equipment failure (n= 3). To recruit participants, letters and emails were sent to families listed in the local newspaper's birth announcements. For their participation, caregivers received an infant t-shirt, bib, or children's book.

#### *Apparatus*

Both the play session and the playback session took place in a 3.7-m x 5.5-m playroom containing infant toys, a toy box, and pictures of animals on the walls. During the play session, infants could move around freely to explore and the size of the room allowed infants to play without continuous interaction with their caregiver. Dyads were video recorded from one of three remote-controlled cameras mounted on the walls. To obtain accurate and detailed recordings of the infant's vocalizations, the infant wore a wireless microphone (Telex FLM-22; Telex Communications, Inc., Burnsville, MN) and transmitter (Telex USR-100) concealed inside the lining of adjustable denim overalls. To obtain accurate recordings of the caregiver speech, mothers also wore a wireless label microphone and transmitter in a pouch around the waist.

For the playback portion, mothers were seated in front of a computer monitor attached to a Wacon ArtZII (12 in x 12 in) tablet. On the 12 in surface, we created a 10 in x 3 in overlay with a vertically-oriented “proximity” response scale that had “Stay where you are” and “Go to baby” as endpoints (Figure 4.1). The tablet was oriented on an angle such that the “Go to baby” endpoint was closest to the monitor. A wireless stylus allowed mothers to tap on the tablet to indicate a response on the ‘proximity’ scale and advance to the next stimulus clip. Mothers listened to each vocalization over headphones (Dynamic Stereo Headphones MDR-7506, SONY). Prior to beginning the playback portion, all of the toys from the play session were placed back in the toy box and a colorful play mat was placed 12 feet away from the mother.



*Figure 4.1.* Playback paradigm. Participants sat in a chair in front of a monitor attached to the tablet. The tablet’s surface was covered with a 10 in x 3 in overlay with a vertically-oriented “proximity” response scale that had “Stay where you are” and “Go to baby” as endpoints. Participants viewed stimulus clips of the infants on the computer monitor and were asked to imagine the infant was located on the play mat located 12 feet across the room (top-left in image). Using the tablet, participants indicated a distance measure of how close they would move after seeing the stimulus clip. Participants also verbally responded to the stimulus clips.



## *Stimuli*

Digital examples of infant behavior were obtained from parent-infant play sessions collected in a previous study. The stimuli were created from recordings of 20 nine-month-old infants, as they tend to produce a wide range of vocalizations and behavior but few or no words. Each stimulus infant provided two vocalizations as auditory stimuli, matched either on vowel resonance (QR or FR) or CV transition timing (MS or CS) as well as two video clips (Table 1). One video showed the infant looking at an object (Object-Directed: ODV) while the other showed the infant looking off screen (Undirected: UDV). The undirected video clip was selected so that the infant was not looking directly at the camera but rather perpendicular to the camera angle, so it was clear that the infant was not looking outward at the participant. The two vocalizations from each infant were each paired with both video clips to create four stimulus clips from each infant. Stimuli were counterbalanced for infant gender. Each stimulus was 5 seconds in duration and clips were edited to have 2.5 seconds of action on either side of the vocalization. Recombining the audio and video could create the possibility of visible mismatches between articulatory movements and sounds that could influence participants' perception of the stimuli (McGurk & MacDonald, 1976). To eliminate the effects of audio/visual recombination, the video clip was paused for the duration of the vocalization. The mothers of the stimulus infants were not shown in any of the clips. The order of the clips was randomized. An additional 12 practice clips preceded the test stimulus set and were created via the same procedure.

Vowel resonance change	QR-->FR
CV transition timing change	MS--> CS

*Table 4.1. Stimulus Categories*

### *Procedure*

Mother-infant dyads came to the laboratory for a single one-hour session. During the play session, mothers were instructed to play with their infants for 15 minutes as they would at home. In the playback phase, mothers responded to a series of audio-visual examples of infant behavior on a computer monitor. Mothers were asked to respond to the stimuli verbally as if they were actually interacting with each stimulus infant and via a computer-driven digitizing tablet and stylus, positioned just below the monitor (Figure 1). Mothers were told to envision that the infant on the video was positioned on a play mat located 12 feet away. For each stimulus, mothers indicated their desired proximity to the infant by touching the tablet with a stylus. Mothers also provided a vocal response to the stimulus clips they felt inclined to respond to.

Mothers were told that the study was about parents' reactions to infants at play and that no fussing or crying infants would be shown. An experimenter provided verbal instructions to the mothers and gave them an opportunity to ask questions. Mothers then read the instructions on the screen again before beginning. Mothers first saw 12 practice stimuli followed by 80 test stimuli. The rate of stimulus presentation was participant-controlled. Upon completion of the 92 audio-visual responses (12 practice, 80 test), mothers heard the complete stimulus set, in a randomized order, and were asked to rate each vocalization on a seven-point 'speechiness' scale, defined as how speech-like the vocalization was (1 = least speech-like, 7 = most speech-like). During the playback phase, the participants' infants played with a researcher in another room.

The order of activities was fixed to allow mothers to become familiar with the toys and room during the play session that they would then see in the stimulus clips during the playback portion.

### *Data Coding and Analysis*

*Infant vocalizations in the play session:* To calculate the frequency of infant vocalizations produced in the play session, vocalizations were divided into syllables. Each syllable was comprised of a single vowel (V) or a consonant and a vowel (either CV or VC). Each syllable was classified according to an infraphonological coding system (Oller & Lynch, 1992; Oller, 2000). Fusses, raspberries and vegetative sounds (e.g. coughs) and sounds with oral obstructions (e.g. toys in the mouth) were excluded from analyses. The first author coded 100% of the infant play sessions and a second coder independently coded 33% of the play sessions. Mean reliability was  $r = .90$  (range= .81-.97) for infraphonological patterning, and  $r = .90$  (range= .88-.92) for directedness of vocalizations.

### *Maternal Responses in play session and playback:*

Mothers' vocal and non-vocal responses to infant behavior in both activities were classified into one of six categories (cf Tamis-LeMonda et al., 2001; Table 2). The first author coded 100% of the mothers' responses to vocalizations during the play sessions and a second coder independently coded 20% of the play sessions. Reliability was  $r = .90$  (range= .80-1.0). Three coders independently categorized maternal responding to the stimuli in the playback paradigm and 20% of the data was recoded as a reliability check. Reliability was  $r = .95$  (range= .91-1.0).

<b>Response Type</b>	<b>Definition</b>	<b>Example</b>
Sensitive	Statements or actions directly related to the object the baby is focused on	<i>That's a ball</i>
Placeholder	Words that hold a place in the conversation without providing new information	<i>Uh-huh, I know</i>
Narrative	Statements related to baby's state or actions	<i>You're so big!</i>
Imitation	Duplications of baby's sound	Baby: [ba]; Mom: [ba]
Redirection	Direct attempt to bring attention to a different place	<i>Look at this toy instead</i>
Non-Sequitur	Statements unrelated to the infant or current context of the infant's environment	<i>What should we have for dinner?</i>

*Table 4.2. Maternal Response Categories*

To validate the playback paradigm as a meaningful method for assessing caregiver responsiveness, we compared mothers' responses to their own infants during the play session with their responses to the playback stimuli. First, we calculated the proportion of vocalizations that received a vocal response during playback. In the playback paradigm, each stimulus clip contained a single infant vocalization (e.g. [ba]), thus caregivers could respond to a maximum of 80 vocalizations. However, in the play session, infants frequently produced many vocalizations in rapid succession (e.g. [ababa]), making it impossible for mothers to respond to all of the individual vocalizations; thus they would look less responsive. To equate the proportion of mother's responding during the play session to responsiveness in the playback paradigm, we calculated the number of phrases that occurred in the play session. A phrase consisted of any vocalizations that occurred within 1 second of each other (e.g. [ababa]; cf Oller & Lynch, 1992). Responsiveness in the play session was calculated as a proportion of the number of responses to the number of vocalization phrases produced; playback responsiveness was calculated as the proportion of vocal responses given out of the total number of vocalizations. Mothers' overall responsiveness and proportion of responses in each response category were compared.

## Results

### *Validation of Playback Methodology*

To assess the degree of similarity between mothers' responses to infant vocalizations in each activity we first conducted a 2 (Activity; Play session, Playback) x 2 (Parity) mixed ANOVA on responsiveness. There were no significant main effect of Activity ( $p = .839$ ) indicating that the proportion of infant phrases mothers' responded to during the play session ( $M = .64, SD = .21$ ) was comparable to the proportion of vocalizations they responded to in the playback paradigm ( $M = .65, SD = .24$ ) (Figure 4.2). There were no other significant main effects or interactions ( $ps > .787$ ).

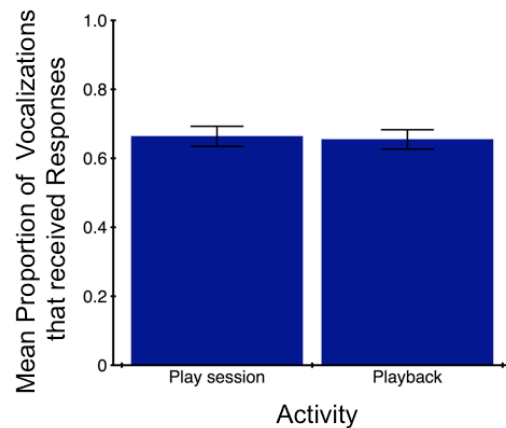
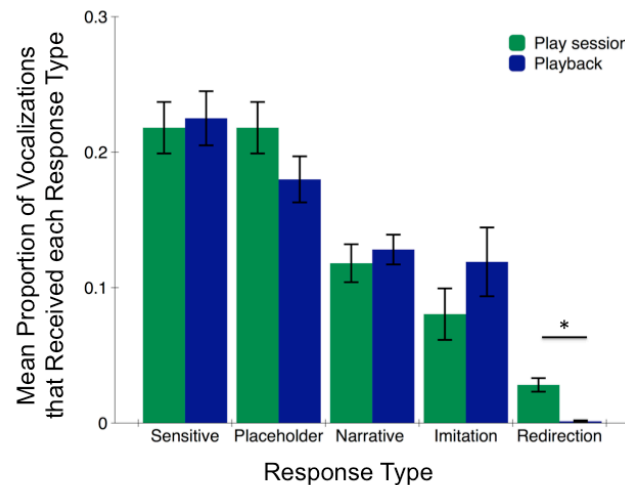


Figure 4.2. Mean proportion of responses by Activity ( $\pm 1$  SE). Mothers' proportions of responses to infant vocalizations between the play session ( $M = .64, SD = .21$ ) and playback ( $M = .65, SD = .24$ ) did not significantly differ.

We further assessed mothers' response profiles on each activity in a 2 (Activity; Play session, Playback) x 6 (Response Type) x 2 (Parity) mixed ANOVA on responsiveness. There was a significant main effect of Response Type,  $F(5, 195) = 46.538, p < .001, \eta_p^2 = .544$ . This main effect was qualified by a significant Activity x Response Type interaction,  $F(5, 195) =$

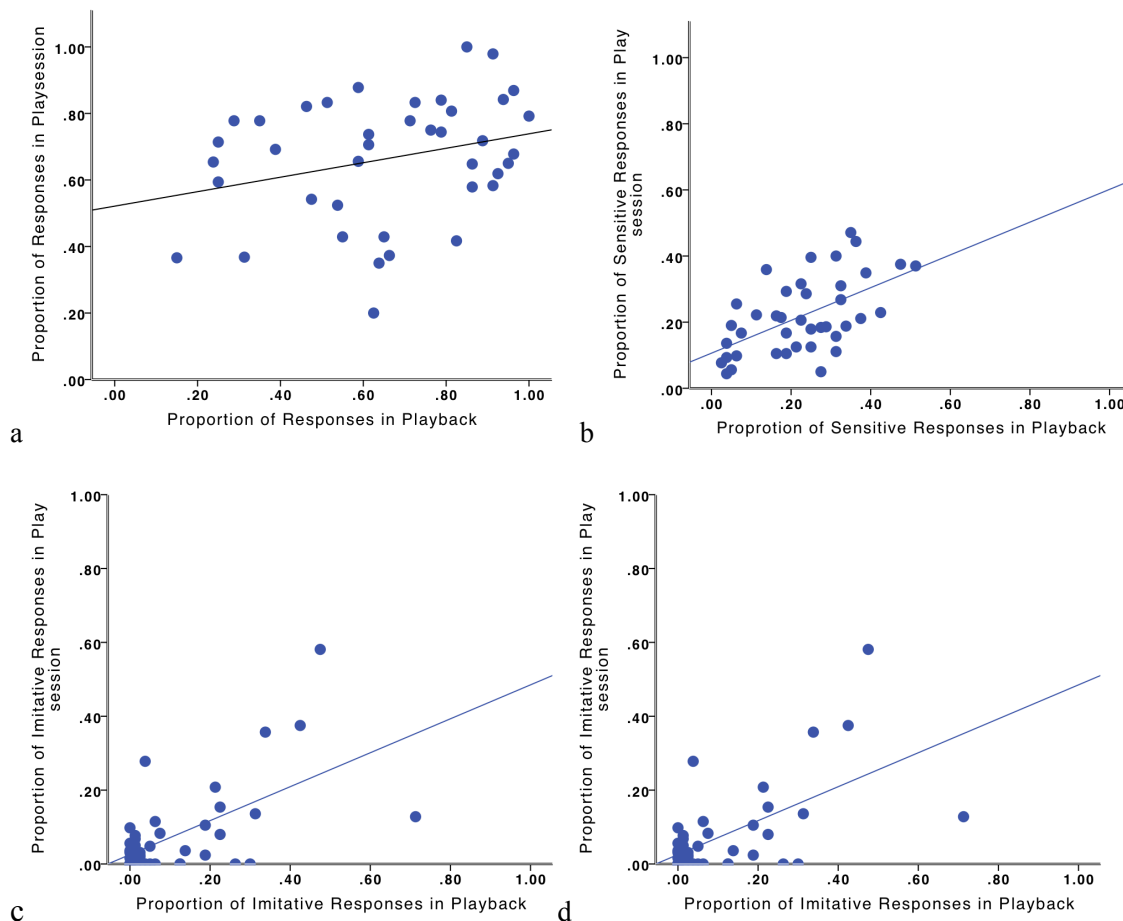
3.117,  $p = .010$ ,  $\eta_p^2 = .074$ . The interaction was decomposed by Response Type (Figure 4.3).

Mothers did not significantly differ in their proportions of Sensitive ( $p = .657$ ), or Narrative ( $p = .60$ ) responses between the two activities. There was a trend for mothers to respond with more placeholders during the play session ( $p = .058$ ), and with more imitative responses ( $p = .068$ ) on the playback. Mothers made significantly more redirective responses during the play session than to the playback stimuli,  $F(1, 39) = 25.113$ ,  $p < .001$ ,  $\eta_p^2 = .392$ . Most redirective responses in the play session resulted from the mother non-verbally shifting the infant's attention to a new object. However, redirective responses to the play back stimuli were difficult to make because mothers did not have access to other objects with which to attract infant attention. As redirective responses were so infrequent in the play back paradigm (<1%), this category was excluded from further analyses. There were also too few responses categorized as comment non-sequiturs in either activity (< .025%) to be analyzed and this category was thus excluded from further analyses.



*Figure 4.3.* Mean proportion of responses to infant vocalizations by Activity and Response Type ( $\pm 1$  SE). Mothers showed similar patterns of responding between the play session and playback. There was a trend for mothers to respond with more Placeholders during the play session ( $p = .058$ ), and with more Imitative responses ( $p = .068$ ) on the playback. Mothers responded with significantly more redirective responses during the play session than during the playback,  $p < .001$ .

As an additional validity check, we correlated mothers' response types on the two activities for the overall proportion of responsiveness to vocalizations and for each maternal response category. Mothers' overall responsiveness was significantly correlated across activities when controlling for the differences in redirective responses ( $r(36) = .330, p = .043$ ; Figure 4.4a). Mothers' sensitive ( $r(38) = .562, p < .001$ ), placeholder ( $r(38) = .430, p = .006$ ) and imitative responses ( $r(38) = .596, p < .001$ ) were also significantly correlated between the two activities (Figures 4.4b, c, d). Mothers' narrative ( $p = .8$ ) and redirective responses ( $p = .9$ ) were not significantly correlated between the two activities.



*Figure 4.4.* Relationship between response types during the play session and playback. Correlations between A: overall proportion of responses between the two activities; B: proportion of sensitive responses; C: proportion of placeholder responses; D: proportion of imitative responses.

## **Discussion**

The results of Experiment 1 indicate that the playback paradigm is a valid method for eliciting natural responses from mothers. Mothers responded similarly to vocalizations whether produced live by their own familiar infant in the play session or displayed over video by unfamiliar infants in the playback paradigm. Mothers did not significantly differ in proportion of overall vocalizations to which they responded between the two activities. They also used sensitive and narrative responses to similar degrees. The only significant difference in response type between the two activities was that mothers provided more redirective responses in the play session. This difference can be attributed to opportunities in the play session to redirect the infant's attention to other objects. Further, mothers' sensitive, placeholder, and imitative responses between the two activities were highly correlated, suggesting that mothers applied their personal response tendencies from interacting with their own infant to the video displays of unfamiliar infants in the playback paradigm. It is unlikely that the similarities in mothers' responses are due to demand characteristics because previous studies demonstrate that parents behave similarly whether interacting in a laboratory or at home (Rothbaum & Crockenberg, 1995). Further, both activities provided warm-up time for participants to become comfortable with the tasks and only the last ten minutes of the play session was coded. Taken together, the playback paradigm is a valid indicator of natural maternal responding.

## **Experiment 2**

Experiment 2 used the playback paradigm to specify the acoustic determinants of maternal responding and investigate how caregiving experience (measured by parity) influences responding to prelinguistic infant vocalizations. We compared the responses of nulliparous, primiparous, and multiparous women.



## Method

### *Participants*

Forty nulliparous undergraduate women ( $M = 19.55$  years; range = 18-22 years) participated for course credit. An additional 18 nulliparous women were tested but excluded for failure to follow instructions ( $n = 11$ ) or equipment failure ( $n = 7$ ). Their responses were compared to the 20 primiparous and 20 multiparous women included in Experiment 1.

### *Apparatus/Stimuli*

The apparatus and stimuli were identical to Experiment 1.

### *Procedure*

Women participated in the playback paradigm and were told that the study was about caregivers reactions to infants at play. Women were told that no fussing or crying infants would be shown and that they should react to the video examples as if they were actually in the play room interacting with the infants. After completing the playback task, participants completed a demographic questionnaire that also included questions assessing their previous caregiving experience. The procedure was otherwise identical to Experiment 1.

### *Coding/ Data Analysis*

Nulliparous women's vocal responses to the play back stimuli were coded using the same categorization system as Experiment 1. Two coders independently categorized participants' responses to the stimuli and 20% of the data was recoded as a reliability check. Reliability was  $r = .943$  (range = .92-1.0).

## Results

### *Perception of Speech-like qualities of stimuli*

To assess women's perceptions of the speech-like quality of infant vocalizations we conducted a 4 (Syllable Type: FRV, QRV, MS, CS) X 3 (Parity) mixed ANOVA on mean speechiness ranking. There was a significant main effect of Syllable Type,  $F(3, 231) = 112.026$ ,  $p < .001$ ,  $\eta_p^2 = .592$ . There was also a significant main effect of Parity,  $F(2, 77) = 6.70$ ,  $p = .002$ ,  $\eta_p^2 = .148$ . These main effects were qualified by a significant Parity x Syllable Type interaction,  $F(6, 231) = 2.534$ ,  $p = .021$ ,  $\eta_p^2 = .062$  (Figure 4.5). The interaction was decomposed by Syllable Type. There was a significant effect of Parity for canonical syllables,  $F(2, 77) = 4.351$ ,  $p = .016$ ,  $\eta_p^2 = .102$ . Tukey's post-hoc tests revealed that nulliparous women rated canonical syllables as less speech-like than did multiparous ( $p = .004$ ) and primiparous ( $p = .037$ ) mothers. There was also a significant effect of Parity for marginal syllables,  $F(2, 77) = 11.77$ ,  $p < .001$ ,  $\eta_p^2 = .234$ . Tukey's post-hoc tests revealed that nulliparous women rated marginal syllables as less speech-like than did multiparous mothers ( $p = .017$ ). For quasi-resonant vowels there was a marginal effect of parity,  $F(2, 77) = 2.825$ ,  $p = .065$ ,  $\eta_p^2 = .068$ . Nulliparous women tended to rate quasi-resonant vowels as less speech-like than did mothers. There were no significant differences in parity for women's perceptions of fully-resonant vowels ( $p = .813$ ).

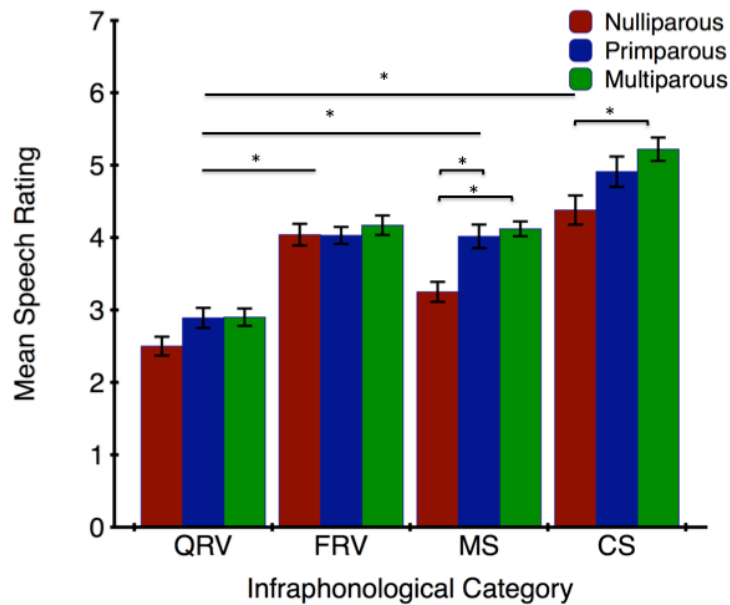


Figure 4.5. Mean ‘speechiness’ rating for each infraphonological type by parity group ( $\pm 1$  SE). Between parity groups, nulliparous women rated marginal syllables as significantly less speech-like than multiparous ( $p = .004$ ) and primiparous ( $p = .037$ ) mothers. Nulliparous women also rated canonical syllables significantly less speech-like than multiparous women ( $p = .017$ ). All three parity groups rated quasi-resonant vowels as less speech-like than the other three infraphonological types ( $ps < .001$ ). Further, nulliparous and multiparous women rated canonical syllables as more speech-like than marginal syllables ( $p < .001$ ), while primiparous women showed a trend to rate canonical syllables as more speech-like than marginal syllables. Further, primiparous and multiparous mothers rated fully-resonant vowels similarly to marginal syllables, while nulliparous women rated fully-resonant vowels similarly to canonical syllables.

The Parity x Syllable Type interaction was also decomposed by parity to further examine the influences of experience on perception of vocal quality. For nulliparous women there was a significant effect of syllable type,  $F(3, 117) = 56.061, p < .001, \eta_p^2 = .585$ . Tukey’s post-hoc tests revealed that nulliparous women perceived quasi-resonant vowels as less speech-like than the other three infraphonological syllable types ( $ps < .001$ ). Further, nulliparous women perceived canonical syllables as more speech-like than marginal syllables ( $p < .001$ ). However, nulliparous women did not perceive a difference in the speech-like quality of canonical syllables and fully-resonant vowels ( $p = .240$ ). For primiparous women there was a significant effect of syllable type,  $F(3, 57) = 25.345, p < .001, \eta_p^2 = .572$ . Tukey’s post-hoc tests revealed that

primiparous women perceived quasi-resonant vowels as less speech-like than the other three infraphonological syllable types ( $p < .001$ ). Further, primiparous women perceived canonical syllables as more speech-like than fully-resonant vowels ( $p = .022$ ) and there was a trend for primiparous women to rate canonical syllables as more speech-like than marginal syllables ( $p = .067$ ). However, primiparous women did not perceive a significant difference in speech-like quality between marginal syllables and fully-resonant vowels ( $p = 1.0$ ). Finally, there was also a significant effect of syllable type for multiparous women,  $F(3, 57) = 56.192, p < .001, \eta_p^2 = .747$ . Tukey's post-hoc tests revealed that multiparous women perceived quasi-resonant vowels as significantly less speech-like than the other three infraphonological syllable types ( $p < .001$ ) and canonical syllables as significantly more speech-like than the other three infraphonological syllable types ( $p < .001$ ). However, multiparous women did not perceive a difference in speech-like quality between marginal syllables and fully-resonant vowels ( $p = 1.0$ ).

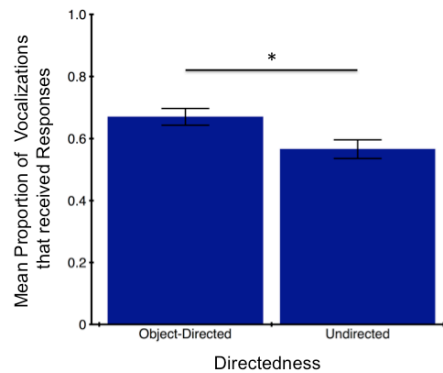
### ***Vocal Responses to Playback Stimuli***

#### *Overall Proportions of Vocal Responses*

To assess the degree of similarity between women's responses to infant vocalizations we conducted a one-way ANOVA to compare the effects of parity on the proportion of playback stimuli that received a verbal response. There was no significant effect of Parity,  $p = .595$ . Nulliparous women ( $M = .59, SD = .25$ ), primiparous women ( $M = .65, SD = .23$ ), and multiparous women ( $M = .65, SD = .25$ ), did not significantly differ in the proportion of playback stimuli that elicited a vocal response.

### *Vocal Responses to Vowel Stimuli*

To assess differences in women's responses to vowels, we conducted a 2 (Directedness: ODV, UDV) x 2 (Vowel Type: FRV, QRV) x 4 (Response Type: Sensitive, Placeholders, Narratives, Imitations) x 3 (Parity: Nulliparous, Primiparous, Multiparous) mixed ANOVA on the proportion of stimuli in the two vowel categories that received a response. There was a significant main effect of Directedness,  $F(1, 77) = 39.699, p < .001, \eta_p^2 = .340$ . Women responded more to vowels when they were in an object-directed context than in an undirected context (Figure 4.6). There was also a significant main effect of Vowel Type,  $F(1, 77) = 27.392, p < .001, \eta_p^2 = .262$ . Women responded to a higher proportion of fully-resonant vowels than quasi-resonant vowels (Figure 4.7). There was also a significant main effect of response type,  $F(3, 231) = 25.653, p < .001, \eta_p^2 = .250$ . Sensitive responses were more frequent than narratives ( $p < .001$ ) and imitations ( $p < .001$ ). Placeholder responses were also more frequent than narratives ( $p < .001$ ) and imitations ( $p < .001$ ).



*Figure 4.6.* Mean proportion of responses to vowel stimuli by directedness of the vocalization ( $\pm 1$  SE). Women responded to a significantly higher proportion of object-directed vowel stimuli than undirected vowel stimuli ( $p < .001$ ).

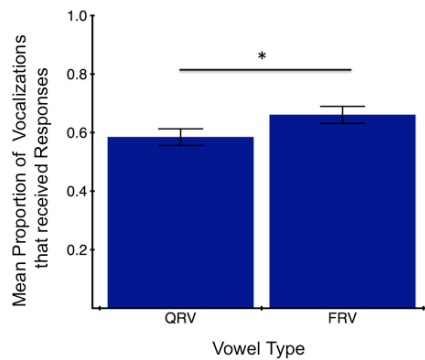


Figure 4.7. Mean proportion of responses to vowel by vocal resonance of the vocalization ( $\pm 1$  SE). Women responded to a significantly higher proportion of fully-resonant vowel stimuli than quasi-resonant vowel stimuli ( $p < .001$ ).

These main effects were qualified by a significant Directedness x Response Type interaction,  $F(3, 231) = 55.832, p < .001, \eta_p^2 = .420$ . The interaction was decomposed by Response Type (Figure 4.8). Women were more likely to respond sensitively to ODVs than UDV's,  $F(1, 79) = 130.371, p < .001, \eta_p^2 = .623$ . Further, women were more likely to respond with placeholders following UDV's than ODVs,  $F(1, 79) = 8.196, p = .005, \eta_p^2 = .094$ . There were also significantly more narrative responses following UDV's than ODVs,  $F(1, 79) = 12.325, p = .001, \eta_p^2 = .135$ . However, the directedness of the vocalization did not significantly influence imitative responses,  $p = .103$ . There was a trend towards a significant Directedness x Vowel Type interaction,  $F(1, 77) = 3.509, p = .065, \eta_p^2 = .044$ . Finally, there was a trend towards a significant Directedness x Vowel Type x Response Type x Parity interaction,  $F(6, 231) = 1.952, p = .074, \eta_p^2 = .048$ . There were no other significant main effects or interactions,  $ps > .212$ .

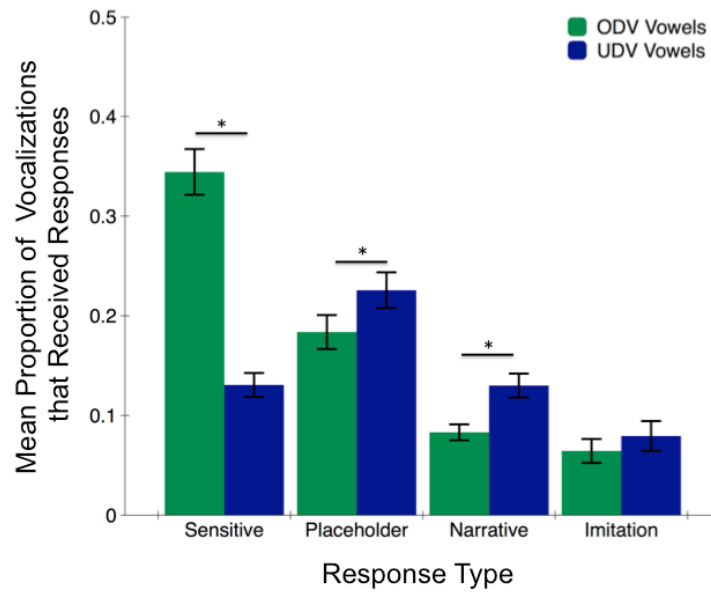


Figure 4.8. Mean proportion of responses to vowel stimuli by directedness of the vocalization and response type ( $\pm$  1 SE). Women responded sensitively to object-directed vowel stimuli significantly more than to undirected vowel stimuli ( $p < .001$ ). However, they responded with both placeholders ( $p = .005$ ) and narratives ( $p = .001$ ) to undirected vowel stimuli significantly more than to object-directed vowel stimuli.

#### *Vocal Latency to Respond to Vowel Stimuli*

To assess differences in latency of women's responses to vowels, we conducted a 2 (Directedness: ODV, UDV)  $\times$  2 (Vowel Type: FRV, QRV)  $\times$  3 (Parity) mixed ANOVA on the mean latency to respond vocally to each of the two vowel categories. There was a significant main effect of Directedness,  $F(1, 72) = 48.069, p < .001, \eta_p^2 = .40$ . Women responded faster to vowels when they were in an object-directed context than in an undirected context (Figure 4.9). Finally, there was a trend towards a significant Directedness  $\times$  Parity interaction,  $F(2, 72) = 2.685, p = .075, \eta_p^2 = .069$ . There were no other significant main effects or interactions,  $ps > .090$ .

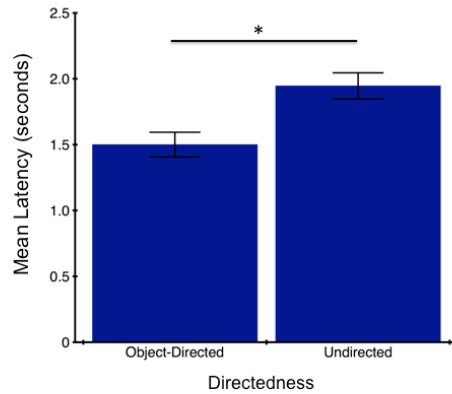


Figure 4.9. Mean latency to produce a vocal response to vowel stimuli by directedness ( $\pm 1$  SE). Women produced a vocal response to object-directed vowel stimuli significantly faster than to undirected vowel stimuli ( $p < .001$ ).

#### *Vocal Responses to Consonant-Vowel Stimuli*

To assess differences in women's responses to CV syllables, we conducted a 2 (Directedness: ODV, UDV)  $\times$  2 (Syllable Type: MS, CS)  $\times$  4 (Response Type: Sensitive, Placeholders, Narratives, Imitations)  $\times$  3 (Parity) mixed ANOVA on the proportion of stimuli from the two CV categories that received a response. There was a significant main effect of Directedness,  $F(1, 77) = 20.394, p < .001, \eta_p^2 = .209$ . Women responded more to CV syllables when they were in an object-directed context than in an undirected context. There was also a significant effect of Response Type,  $F(3, 231) = 6.201, p < .001, \eta_p^2 = .075$ . Women produced significantly fewer imitations than sensitive ( $p = .007$ ) and placeholder ( $p = .006$ ) responses. These main effects were qualified by a significant Directedness  $\times$  Syllable type interaction,  $F(1, 77) = 4.361, p = .040, \eta_p^2 = .054$  (Figure 4.10). The interaction was decomposed by Directedness. For undirected vocalizations, women responded significantly more to marginal syllables than to canonical syllables,  $F(1, 79) = 4.348, p = .040, \eta_p^2 = .052$ . There was no significant effect of syllable type for object-directed vocalizations,  $p = 1.0$ .



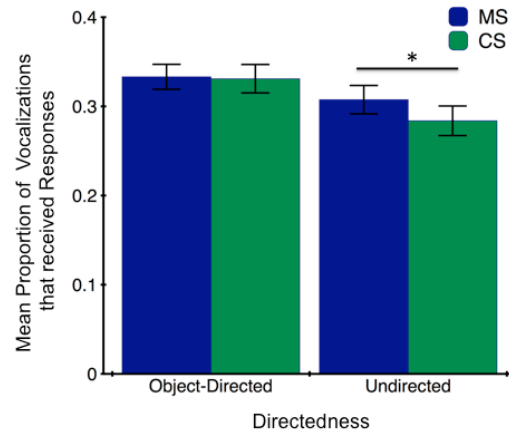


Figure 4.10. Mean proportion of responses to consonant-vowel stimuli by directedness of the vocalization and consonant-vowel type ( $\pm 1$  SE). Women responded significantly more to undirected marginal syllables than undirected canonical syllables ( $p = .040$ ).

There was also a significant Directedness  $\times$  Response Type interaction,  $F(3, 231) = 39.310, p < .001, \eta_p^2 = .338$  (Figure 4.11). The interaction was decomposed by Response Type. Women were more likely to respond sensitively to ODVs than to UDVs,  $F(1, 79) = 87.503, p < .001, \eta_p^2 = .526$ . In contrast, women were more likely to respond with narratives following UDVs than to ODVs,  $F(1, 79) = 36.341, p < .001, \eta_p^2 = .315$ . However, directedness of the vocalization did not influence mothers' likelihood of responding with placeholders ( $p = .731$ ) or imitations ( $p = .211$ ).

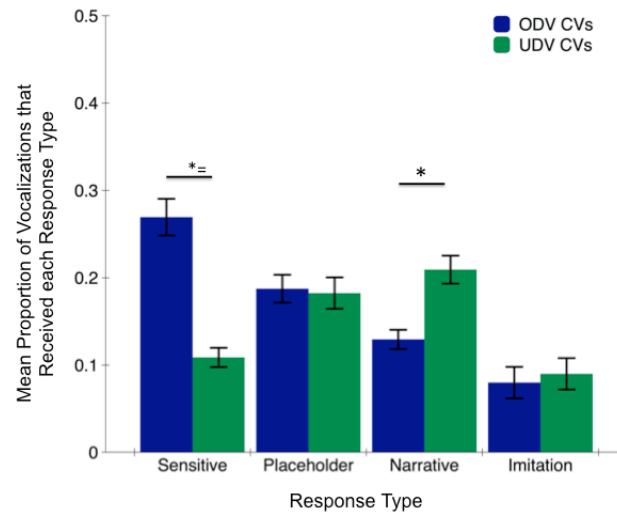


Figure 4.11. Mean proportion of responses to consonant-vowel stimuli by directedness of the vocalization and response type ( $\pm 1$  SE). Women responded sensitively to object-directed consonant-vowels significantly more than to undirected consonant-vowel stimuli ( $p < .001$ ). However, they responded with narratives to undirected consonant-vowel stimuli significantly more than to object-directed consonant-vowel stimuli ( $p < .001$ ).

There was a significant Syllable Type x Response Type interaction,  $F(3, 231) = 12.290$ ,  $p < .001$ ,  $\eta_p^2 = .139$  (Figure 4.12). The interaction was decomposed by Response Type. Women responded with narratives more often to marginal syllables than to canonical syllables,  $F(1, 79) = 20.354$ ,  $p < .001$ ,  $\eta_p^2 = .205$ . However, women responded by imitating canonical syllables more than marginal syllables,  $F(1, 79) = 19.735$ ,  $p < .001$ ,  $\eta_p^2 = .20$ . There was also a marginal effect of Syllable Type for placeholder responses,  $F(1, 79) = 3.695$ ,  $p = .058$ ,  $\eta_p^2 = .045$ . Women tended to respond with placeholders more in response to marginal syllables than to canonical syllables. Syllable Type did not significantly influence mothers' sensitive responses ( $p = .20$ ).

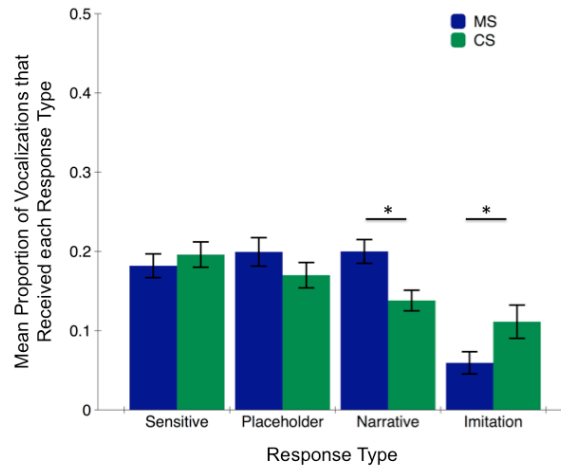


Figure 4.12. Mean proportion of responses to consonant-vowel stimuli by consonant-vowel type and response type ( $\pm 1$  SE). Women responded with narratives significantly more to marginal syllables than canonical syllables ( $p < .001$ ). However, women imitated canonical syllables significantly more than marginal syllables ( $p < .001$ ).

Finally, there was a marginally significant Response Type x Parity interaction,  $F(6, 231) = 2.111$ ,  $p = .053$ ,  $\eta_p^2 = .052$  (Figure 4.13). Planned comparisons were conducted to assess differences in response types between parity groups. There was a significant effect of parity for imitative responses,  $F(1, 77) = 4.358$ ,  $p = .015$ ,  $\eta_p^2 = .102$ . Post-hoc Tukey's HSD tests revealed that Nulliparous women were less likely to imitate than Multiparous women ( $p = .036$ ) and tended to imitate less than Primiparous women ( $p = .063$ ). There were no significant effects of parity for the other three response types,  $ps > .461$ . There were no other significant main effects or interactions,  $ps > .100$ .

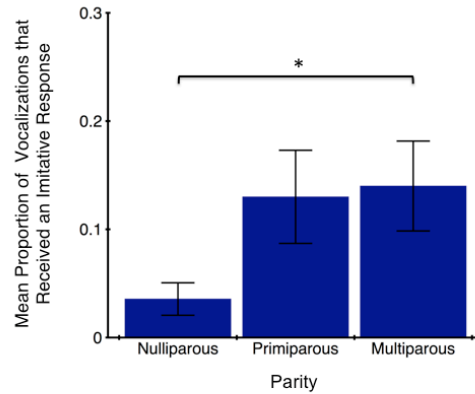


Figure 4.13. Mean proportion of imitative responses to consonant-vowel stimuli by parity group ( $\pm 1$  SE). Nulliparous women were significantly less likely to imitate consonant-vowel stimuli than multiparous mothers,  $p = .036$ .

#### *Vocal Latency to Respond to Consonant-Vowel Stimuli*

To assess the differences in latency of women's responses to consonant-vowels, we conducted a 2 (Directedness: ODV, UDV)  $\times$  2 (CV type: MS, CS)  $\times$  2 (Parity) mixed ANOVA on the mean latency to respond vocally to the two CV categories. There were no significant main effects or interactions,  $ps > .098$ .

#### *Vocal Responses to Infant Phonological Types*

To assess the impacts of phonological type (vowel vs. consonant-vowel vocalizations) on women's responses, we conducted a 2(Directedness: ODV, UDV)  $\times$  2(Phonology: V, CV)  $\times$  4 (Response Type)  $\times$  3 (Parity) mixed ANOVA on mothers' proportions of responses to each stimulus type. There was a significant main effect of Directedness,  $F(1, 77) = 47.871$ ,  $p < .001$ ,  $\eta_p^2 = .383$ . There was also a significant main effect of Response Type,  $F(3, 231) = 14.944$ ,  $p < .001$ ,  $\eta_p^2 = .163$ . These main effects were qualified by significant interactions of Directedness  $\times$  Phonology ( $F(1, 77) = 4.639$ ,  $p = .034$ ,  $\eta_p^2 = .057$ ), Directedness  $\times$  Response

Type ( $F(3, 231) = 65.159, p < .001, \eta_p^2 = .458$ ), and Phonology x Response Type ( $F(3, 231) = 16.427, p < .001, \eta_p^2 = .176$ ).

Finally, there was a significant Directedness x Phonology x Response Type interaction,  $F(3, 231) = 6.496, p = .001, \eta_p^2 = .078$ . The interaction was decomposed by Response Type. For Sensitive responses, there was a significant Directedness x Phonology interaction,  $F(1, 79) = 8.568, p = .004, \eta_p^2 = .098$  (Figure 4.14a). For vocalizations in an object-directed context, women responded sensitively more often to vowel vocalizations than consonant-vowel vocalizations,  $F(1, 79) = 21.224, p < .001, \eta_p^2 = .212$ . There was also a significant effect of Phonology for undirected vocalizations,  $F(1, 79) = 4.234, p = .043, \eta_p^2 = .051$ . Women responded sensitively more often to undirected vowel vocalizations than to undirected consonant-vowel vocalizations.

For placeholders, there was a significant Directedness x Phonology interaction,  $F(1, 79) = 8.184, p = .005, \eta_p^2 = .094$  (Figure 4.14b). When vocalizations were undirected, women were more likely to respond with a placeholder to Vowel sounds than to CV sounds,  $F(1, 79) = 10.277, p = .002, \eta_p^2 = .115$ . There was no significant effect of Phonology for object-directed vocalizations,  $p = .781$ .

For narrative responses, there was a significant Directedness x Phonology interaction,  $F(1, 79) = 4.578, p = .035, \eta_p^2 = .055$  (Figure 4.14c). Women responded with narratives more to consonant-vowel vocalizations than to vowel vocalizations for both ODVs ( $F(1, 79) = 17.698, p < .001, \eta_p^2 = .183$ ) and UDV's ( $F(1, 79) = 28.473, p < .001, \eta_p^2 = .265$ ). There was no significant Directedness x Phonology interaction for imitative responses ( $p = .675$ ). There were no other significant main effects or interactions,  $ps > .270$ .

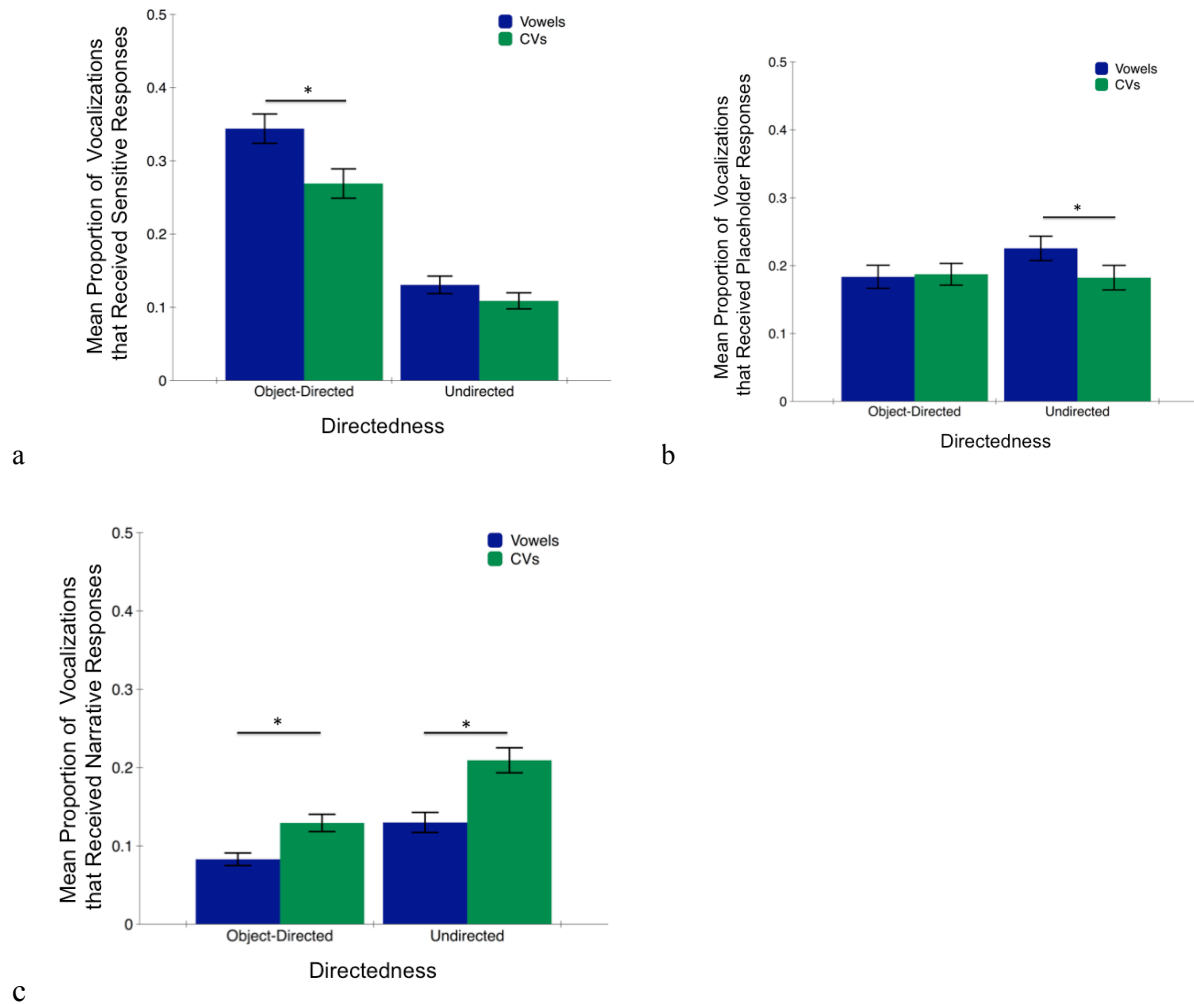


Figure 4.14. Mean proportion of responses to phonological types by directedness of the vocalization and phonology type ( $\pm 1$  SE). A. Sensitive responses: Women were significantly more likely to respond sensitively to vowels than consonant-vowels for both object-directed and undirected stimuli ( $p < .001$ ); B. Placeholder responses: When vocalizations were undirected, women were significantly more likely to respond with placeholders to vowels than consonant-vowels ( $p = .002$ ); C. Narratives: Women were significantly more likely to respond with narratives to consonant-vowels than vowels for both object-directed and undirected stimuli ( $p < .001$ ).

### *Vocal Latency Differences for Phonological Types*

To assess the impacts of phonological type (vowel vs. consonant-vowel vocalizations) on women's latencies to respond we conducted a 2 (Directedness: ODV, UDV)  $\times$  2 (Phonology: V, CV)  $\times$  3 (Parity) mixed ANOVA on the mean latency to indicate a proximity response for each stimulus category. There was a significant main effect of Directedness,  $F(1, 69) = 31.362$ ,

$p < .001$ ,  $\eta_p^2 = .312$ . There was also a significant main effect of Phonology,  $F(1, 69) = 11.103$ ,  $p = .001$ ,  $\eta_p^2 = .139$ . These main effects were qualified by a significant Directedness x Phonology interaction,  $F(1, 69) = 10.803$ ,  $p = .002$ ,  $\eta_p^2 = .086$ .

Further, there was a significant Directedness x Phonology x Parity interaction,  $F(2, 69) = 3.245$ ,  $p = .045$ ,  $\eta_p^2 = .086$  (Figure 4.15). The interaction was decomposed by Parity. For Primiparous mothers, there was a significant Directedness x Phonology interaction,  $F(1, 17) = 13.673$ ,  $p = .002$ ,  $\eta_p^2 = .446$ . This interaction was decomposed by Directedness. There was a significant effect of Phonology for ODVs,  $F(1, 19) = 14.332$ ,  $p = .001$ ,  $\eta_p^2 = .430$ . Primiparous women responded faster to object-directed vowels than to object-directed consonant-vowels. There was no significant effect of Phonology for UDVs,  $p = .459$ . There were no significant Directedness x Phonology interactions for Nulliparous ( $p = .261$ ) or Multiparous ( $p = .515$ ) women. There were no other significant main effects or interactions,  $ps > .254$ .

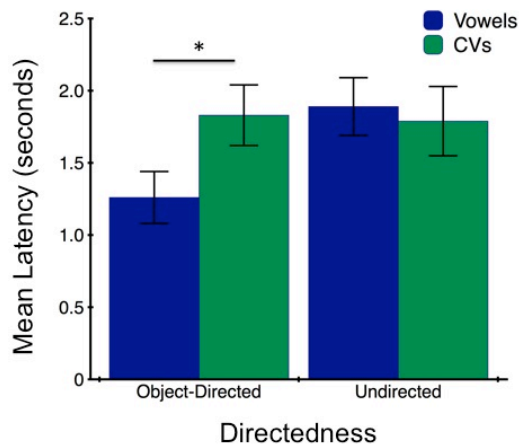
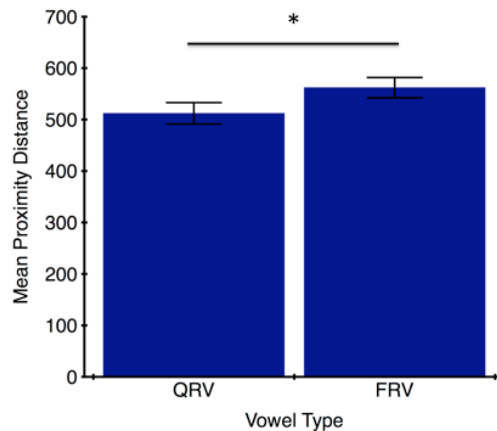


Figure 4.15. Mean latency of primiparous mothers to produce a vocal response by directedness and phonological type ( $\pm 1$  SE). Primiparous mothers produced a vocal response to object-directed vowel stimuli significantly faster than to object-directed consonant-vowel stimuli ( $p < .001$ ).

### ***Proximity Measure Responses***

*Mean Proximity Responses to Vowels:* To assess differences in women's responses to vowels using the proximity scale, we conducted a 2 (Directedness: ODV, UDV) x 2 (Vowel Type: QRV, FRV) x 3 (Parity) mixed ANOVA on the mean distance response chosen for each vowel category. There was a significant main effect of Vowel type,  $F(1, 74) = 8.059, p = .006, \eta_p^2 = .098$ . Women indicated that they would move closer to the infant using the proximity measure in response to fully-resonant vowels (Figure 4.16). Finally, there was a trend towards a significant Directedness x Parity interaction,  $F(2, 74) = 2.720, p = .072, \eta_p^2 = .068$ . There were no other significant main effects or interactions,  $ps > .094$ .



*Figure 4.16.* Mean proximity response by vowel type ( $\pm 1$  SE). Women indicated that they would move significantly closer to the infant after hearing a fully-resonant vowel than a quasi-resonant vowel ( $p = .006$ ).

*Mean Proximity Responses to Consonant-Vowels:* To assess differences in women's responses to consonant-vowels using the proximity scale, we conducted a 2 (Directedness: ODV,



UDV) x 2 (CV type: MS, CS) x 3 (Parity) mixed ANOVA on the mean distance response indicated for each CV category. There were no significant main effects or interactions,  $ps > .090$ .

*Mean Proximity Responses by Phonological Type:* To assess the impacts of Phonological type (vowel vs. consonant-vowel vocalizations) on women's proximity responses, we conducted a 2 (Directedness: ODV, UDV) x 2 (Phonology: V, CV) x 3 (Parity) mixed ANOVA on the mean distance response indicated for each stimulus category. There was a significant main effect of Phonology,  $F(1, 75) = 17.145, p < .001, \eta_p^2 = .186$ . Women indicated that they would move closer to the infant using the proximity measure after hearing a vowel stimulus than a consonant-vowel stimulus. This main effect was qualified by a significant Directedness x Phonology interaction,  $F(1, 75) = 4.991, p = .028, \eta_p^2 = .062$  (Figure 4.17). The interaction was decomposed by Directedness. Women responded that they would move closer to the infant using the proximity measure following undirected vowels,  $F(1, 77) = 23.116, p < .001, \eta_p^2 = .231$ . There was no significant effect of Phonology for object-directed stimuli,  $p = .153$ . Finally, there was a significant main effect of Parity,  $F(2, 75) = 3.165, p = .048, \eta_p^2 = .078$ . However, post-hoc Tukey's HSD tests revealed no significant differences between parity types,  $ps > .102$ . Visual inspection of the data suggests that nulliparous women had lower mean proximity measures than both primiparous and multiparous women, indicating that they would be less likely to move towards the infant in response to the stimuli. There were no other significant main effects or interactions,  $ps > .461$ .

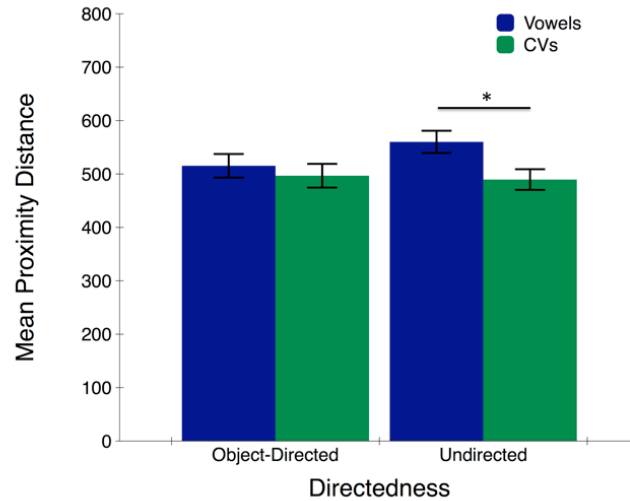


Figure 4.17. Mean proximity response to phonological types by directedness of the vocalization and phonology type ( $\pm 1$  SE). Women indicated that they would move significantly closer to the infant after hearing an undirected vowel than an undirected consonant-vowel ( $p < .001$ ).

### ***Proximity Latency Measures***

*Latency to Respond to Vowels:* To assess the differences in latency of women's responses to vowels using the proximity measure, we conducted a 2 (Directedness: ODV, UDV) x 2 (Vowel type: QRV, FRV) x 3 (Parity) mixed ANOVA on the mean latency to indicate a proximity response for each vowel category. There were no significant main effects or interactions,  $ps > .342$ .

*Latency to Respond to Consonant-Vowels:* To assess the differences in latency of women's responses to consonant-vowels using the proximity measure, we conducted a 2 (Directedness: ODV, UDV) x 2 (CV type: MS, CS) x 3 (Parity) mixed ANOVA on the mean latency to indicate a proximity response for each CV category. There was a trend towards significant Directedness x CV type interaction,  $F(1, 74) = 3.015, p = .087, \eta_p^2 = .039$ . There were no significant main effects or interactions,  $ps > .099$ .

*Comparisons of Latency by Phonological Type:* To assess the impacts of phonological type (vowel versus consonant-vowel vocalizations) on latency of women's responses, we conducted a 2 (Directedness: ODV, UDV) x 2 (Phonology: V, CV) x 3 (Parity) mixed ANOVA on the mean latency to indicate a proximity response for each stimulus category. There were no significant main effects or interactions,  $ps > .193$ .

## **Discussion**

The results of Experiment 2 demonstrate that caregivers were influenced by both the vocal quality and context of prelinguistic vocalizations. With regards to context, women were significantly more likely to vocally respond to object-directed vocalizations than to undirected vocalizations. Women provided faster vocal responses to object-directed vowels than to undirected vowels. The directedness of the vocalization also impacted the types of responses women provided. Regardless of vocal quality, women provided significantly more sensitive responses to object-directed vocalizations than to undirected vocalizations. Women were also significantly more likely to provide narrative responses to undirected vocalizations than to object-directed vocalizations. Finally, women responded with more placeholder responses to undirected vowels than to object-directed vowels. We expected that caregivers would respond sensitively in response to ODVs given that the infant is already attending to an object to comment on. In the same vein, perhaps caregivers provided more narrative and placeholder responses to undirected vocalizations to continue the protoconversation when there was no clear referent of the infant's attention. The differential information content of responses to object-directed and undirected vocalizations may provide infants with different opportunities for learning.

Women also demonstrated perceptual sensitivity to the infraphonological properties of vocalizations. While previous studies have shown that vocal quality influences mothers' ratings of infant attractiveness and mood (Bloom & Lo, 1990), this experiment is the first to demonstrate that mothers perceive differences in the speech-like quality of infant vocalizations based on infraphonological type. In response to vowel stimuli, women responded to a significantly higher proportion of fully-resonant vowels than quasi-resonant vowels. However, women showed similar proportions of response types whether responding to fully-resonant or quasi-resonant vowels. The quality of the vowel thus influenced women's likelihood of responding to the vocalization but not how they responded.

In response to consonant-vowel stimuli, women showed similar proportions of vocal responses between marginal and canonical syllables. However, the type of vocal response they provided was influenced by syllable type. Women were more likely to narrate in response to marginal syllables than canonical syllables, particularly when vocalizations were undirected. Perhaps marginal syllables are perceived as speech-like enough to attract mothers' attention. However, when the vocalization was undirected there was no clear object to react to besides the infant, so we infer that mothers commented on infants' behaviors as a way of taking a conversational turn and continuing the interaction. In contrast, women were more likely to imitate canonical syllables than marginal syllables. Previous work suggests that imitation acts as a 'social glue', demonstrating similarity and social affiliation while also promoting prosocial behavior (van Baaren et al. 2004; Carpenter, Uebel, & Tomasello; 2013; Chartrand & Bargh, 1999). Women may imitate the sounds they perceive as most speech-like in an attempt to increase affiliation with the infant. Surprisingly, there was a trend for nulliparous women to

imitate infant consonant-vowel syllables less than did mothers, suggesting that imitation of infant behavior is a phenomenon that developed with caregiving experience.

When comparing responses to vowel stimuli with responses to consonant-vowel stimuli, there were several notable interactions. First, women were more likely to respond sensitively to object-directed vowels than to object-directed consonant-vowel syllables. Further, women provided more placeholder responses to undirected vowels than to undirected consonant-vowels. Women were also more likely to narrate in response to consonant-vowels than to vowels alone regardless of the directedness of the vocalization. Also, when vocalizations were directed at objects, women were faster to vocally respond to vowels than to consonant-vowel syllables. These results suggest that the directedness of the vocalizations influenced how, when, and the speed at which women reacted to less mature vocalizations (e.g. vowels). In contrast, the directedness of the vocalization was less influential on women's reactions to more advanced sounds.

The current experiment is the first to demonstrate that parity influences perception of the quality of infant vocalizations. All three parity groups perceived quasi-resonant vowels as least speech-like, but perceptions of other three categories were influenced by parity. Nulliparous women rated canonical syllables as less speech-like than did mothers and marginal syllables as less speech-like than did multiparous mothers. Further, while nulliparous women did not distinguish between fully-resonant vowels and canonical syllables, both primiparous and multiparous mothers rated canonical syllables as more speech-like than fully-resonant vowels. Parity also influenced mothers' speech ratings. Multiparous mothers rated canonical syllables as more speech-like than marginal syllables while primiparous mothers did not rate the two

categories differently. Thus, perception of infant vocal quality developed with increasing caregiving experience.

The major differences in responsiveness were on the verbal measures and speech ratings. In contrast, there were minimal differences in caregiver responses using the proximity measure. The only significant finding was that women responded that they would move closer to the infant after hearing fully-resonant vowels than quasi-resonant vowels, further demonstrating that women respond differently to the sounds they perceive as more speech-like. The lack of differences between infraphonological types, directedness, and parity types using the proximity measure was surprising. We developed the proximity measure to approximate caregivers' non-vocal behavior when interacting with infants, but the one-dimensional nature of the distance measure was perhaps too simple to approximate natural caregiver responding. Also, caregivers may have struggled to provide both verbal and proximity responses at the same time. Our latency measures may have been inflated as participants tended to make a verbal response and then look to the tablet to indicate a proximity response. Future studies could attempt to test whether separating the tasks yields different results.

## **General Discussion**

In Experiment 1, we validated the playback paradigm as a new methodology for investigating caregiver responsiveness to infants' prelinguistic vocalizations. While the playback paradigm has been used in a limited fashion to study caregiver perceptions of infant vocalizations and cries (Bloom, D'Odorico, & Beaumont, 1993; Goldstein & West, 1999; Green, Jones, Gustafson, 1987; Papoušek, 1989), the methodology has never been validated as a meaningful indicator of maternal *behavior*. In Experiment 1, we directly compared the same

mothers' responses to their own infant's vocalizations in a play session with their responses to unfamiliar infants' vocalizations using the playback paradigm. The results indicate that mothers' vocal responses to audio-visual examples of unfamiliar infants' vocal behavior follow the same patterns as their reactions to their own infants.

Experiment 2 manipulated the acoustic qualities and directedness of infant vocalizations to systematically examine their effects on women with differing amounts of caregiving experience. Women vocally responded more to vocalizations from an object-directed context than an undirected context. Further, they vocally responded to a higher proportion of fully-resonant vowels than quasi-resonant vowels. Women also indicated they would move closer to the infant in response to fully-resonant vowels than quasi-resonant vowels. Finally, women provided different kinds of vocal responses to consonant-vowel stimuli depending on the infraphonological type.

Our results also provide insight into the development of caregiving behavior. Each of the three parity groups perceived differences in the four infraphonological categories in significantly different ways suggesting that perception of the speech-like qualities of infant vocalizations continuously develops with increasing experience. Given these perceptual differences, it is possible that infant vocalizations attract and organize nulliparous women and primiparous mothers' attention differently than multiparous mothers. Parity also influenced the types of vocal responses women provided. Nulliparous women tended to imitate infant vocalizations less than multiparous women, suggesting that responding to infant vocalizations develops with caregiving experience.

With respect to the development of vocal communication and language, our findings suggest that babbling organizes the social interactions that elicit contingent feedback from

caregivers. Previous work demonstrates that contingent feedback to infants' vocalizations facilitates real-time vocal development and word learning (Goldstein & Schwade, 2008; 2010; Goldstein et al., 2010). However, in the playback paradigm, not all vocalizations were equally likely to elicit a response from caregivers. Caregivers tended to respond more frequently to more speech-like vowels and vocalizations directed at objects. The acoustic qualities of the consonant-vowel vocalizations also influenced the types of responses that caregivers provided. In live interactions with infants, the types of caregiver responses have different effects on later learning (e.g. Akhtar, Dunham & Dunham, 1991; Stevens, Blake, Vitale & MacDonald, 1998). The differential information caregivers provide based on vocal quality and directedness may provide infants with different opportunities for learning. Understanding the qualities of infant vocalizations that reliably received caregiver responses brings us closer to understanding which aspects of moment-to-moment interactions between caregivers and infants facilitate language development.

Having validated the playback paradigm as a meaningful measure of mothers' responsiveness to infant vocalizations, future experiments can utilize this paradigm to further investigate caregiver responses to infant vocalizations and behaviors. There are an unlimited number of infant behaviors that could be manipulated and caregiver qualities that could be controlled for in future studies. For example, in the current experiments participants responded to stimuli consisting of single syllables. However, infants often make multi-syllable utterances and individual vocalizations vary in duration. Future studies will manipulate the length of vocalizations and number of syllables to examine the effects of duration and concatenation on caregiver responding.



A second question to explore is what accounts for the changes in caregiver responsiveness and perception of the speech-like qualities of infant vocalizations with increasing parity? One possibility is that childbirth may reorganize maternal responsiveness in women due to changes in hormones associated with delivery. For example, prolactin, a protein that regulates lactation in females, increases feeding behavior and decreases stress in new mothers (Ma et al., 2005). Injecting prolactin into nulliparous rats reduces the latency to onset of maternal behavior (Bridges et al., 1990), while blocking the signaling of prolactin can delay maternal responding (Bridges et al., 2001). Female rats increase their retrieval responses to pup vocalizations after giving birth, and they continue to increase their responsiveness as they gain more experience with their pups (Farrell & Alberts, 2002). These changes are associated with endocrine changes accompanying pregnancy and birth. In humans, the release of oxytocin and prolactin in childbirth is associated with bonding between the mother and infant following birth (Uvnas-Moberg, 1998). Thus, parturition represents a unique opportunity to use a within-subjects design to study changes in maternal responsiveness to prelinguistic vocalizations. In the future, we will extend these findings by testing nulliparous pregnant women before and after parturition. Such findings could eventually be linked directly to changes in endocrine state that accompany late pregnancy and birth.

An alternative hypothesis for why the different parity groups perceive differences in acoustic quality of prelinguistic vocalizations differences may be that parity serves as a proxy for exposure in our experiment. In other words, the development of caregiver responsiveness could be a function of learning what constitutes a good vocalization. Multiparous women may perceive sounds differently simply because they have heard many more prelinguistic vocalizations than nulliparous and primiparous women as a result of having multiple children. It is possible that

with enough experience or exposure to infant vocalizations, nulliparous women would perceive and respond to vocalizations in the same ways as more experienced women. To test the exposure hypothesis in a future study we will test nulliparous women with extensive infant caregiving experience, such as day-care workers, with our stimuli set using the playback paradigm.

Future work will also investigate how maternal responsiveness changes as infants develop more mature vocalizations. Research by Bornstein and colleagues (2008) demonstrates that maternal responsiveness is dynamic and response types change as infants display more mature behaviors and vocalizations. As infants age caregivers respond less by describing and encouraging exploration of objects and increasingly respond with vocal imitations and questions (Bornstein et al., 2008). Assessing caregivers using playback stimuli when their infants are at different levels of vocal development will also lend insight into how quality of infant vocalizations organize caregiver responses. For example, will a mother of a younger infant, such as a five-month-old who is not yet producing canonical syllables, respond to marginal syllables in the same ways that a mother of a 9-month-old responds to a canonical syllable, or will that mother respond more to canonical syllables in the same way that the mother of a 9-month-old infant would respond? Mothers may scale their responses to the most mature syllable that their infant currently produces, or they may scale their responses to the vocalizations that are most comparable to adult speech. Testing mothers of infants at various stages of vocal development will illuminate infants' contributions to caregiver-infant interactions and how perceptions of the speech-like qualities of infant vocalizations develop.

Another outstanding question is whether infants learn differently from maternal responses to a more or less mature vocalization. Previous research demonstrates that infants who produce more canonical syllables at 1 year of age have more advanced vocabulary and speech later in

development (Stoel-Gammon, 1992). Further, children who have a delay in the onset of canonical babbling tend to be 'late talkers', showing delays in their productive vocabulary development (Paul & Jennings, 1992; Rescorla & Ratner, 1996). While these results were correlational, the results of the playback paradigm provide a possible explanation of the relationship between early babbling and later language. Early vocalizations appear to contribute to a feedback loop in which parents reinforce more speech-like vocalizations, particularly with sensitive responses such as object labels, which may then promote word learning.

Previous work examining infant vocal learning in moment-to-moment interactions demonstrates that infants learn phonological patterns and words from contingent feedback to their vocalizations (Goldstein & Schwade, 2008; Goldstein, Syal, & Schwade, under review; Goldstein et al., 2010). Using a similar vocal learning paradigm, mothers could be cued only to respond with new phonological patterns following immature vocalizations or mature vocalizations. Infants' production of the new phonological form would then be assessed. The effects of learning from responses to mature versus immature vocalizations could also be tested using a word learning paradigm. For example, novel objects could be labeled immediately following infants' ODVs while controlling for the quality of the infant's ODV to assess differential effects of vocal quality on learning word-object associations (cf Goldstein and colleagues, 2010).

Finally, we initially limited our sample to women to validate the playback paradigm and investigate caregiver responsiveness to infant vocalizations. However, paternal responsiveness to infant behavior is relatively understudied (Parke, 2000). We are currently testing fathers and nulliparous males in the playback paradigm to determine the impacts of prelinguistic

vocalizations on paternal responding. We will then compare differences between parity groups and gender.

### **Conclusions**

The present study is a first step in understanding the characteristics of infant vocal behavior that drive maternal responses that then influence infant learning. We introduce the playback paradigm as a validated experimental tool for research on early communicative development. The playback paradigm demonstrated that women's responses to infant babbling are influenced by both the quality and directedness of vocalizations. Caregivers' perceptions of infant vocalizations and responsiveness also develop with increasing caregiving experience. The current experiments therefore highlight the importance of responding to immature sounds in constructing social interactions that facilitate language development.

## References

- Akhtar, N., Dunham, F. & Dunham, P. J. (1991). Directive interactions and early vocabulary development: The role of joint attentional focus. *Journal of Child Language*, 18, 41-49.
- Albert, R. R., Schwade, J. A., & Goldstein, M. H. (April, 2013). *The Influence of Caregiving Experience on Caregiver-Infant Interactions*. Poster presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.
- Baumwell, L., Tamis-LeMonda, C. S. & Bornstein, M.H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development*, 20, 247-258.
- Beaumont, S., & Bloom, K. (1993). Adult attributions of intentionality to vocalizing infants. *First Language*, 13, 235-247.
- Bloom, K., D'Odorico, L., & Beaumont, S. (1993). Adult preferences for syllabic vocalizations: Generalizations to parity and native language. *Infant Behavior and Development*, 16, 109-120.
- Bloom, K. & Lo, E. (1990). Adult perceptions of vocalizing infants. *Infant Behavior and Development*, 12, 209-219.
- Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, 44, 867-874.
- Bornstein, M. H. & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *New Directions in Child Development*, 43, 49-61.

- Bridges, R.S., Numan, M., Ronsheim, P.M., Mann, P.E., & Lupini, C.E. (1990). Central prolactin infusions stimulate maternal behavior in steroid-treated, nulliparous female rats. *Proceedings of the National Academe of Science*, 87, 8003-8007.
- Bridges, R.S, Rigerio, B.A., Byrnes, E.M., Yang, L.L., & Walker, A.M. (2001). Central infusions of the recombinant human prolactin receptor antagonist, S179D-PRL, delay the onset of maternal behavior in steroid-primed, nulliparous female rats. *Endocrinology*, 142, 730-739.
- Carpenter, M., Uebel, J., & Tomasello, M. (2013). Being Mimicked Increases Prosocial Behavior in 18-Month-Old Infants. *Child Development*.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910.
- Crockenberg, S., & Litman, C. (1990). Autonomy as competence in 2-year-olds: Maternal correlates of child defiance, compliance, and self-assertion. *Developmental Psychology*, 26, 961–971.
- Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36, 511-523.
- Della Corte, M., Benedict, H. & Klein, D. (1983). The relationship of pragmatic dimensions of mothers' speech to the referential-expressive distinction. *Journal of Child Language*, 10, 35-43.
- Farrell, W. J., & Alberts, J. R. (2002). Maternal responsiveness to infant Norway rat (*Rattus norvegicus*) ultrasonic vocalizations during the maternal behavior cycle and after steroid and experiential induction regimens. *Journal of Comparative Psychology*, 116, 286-296.

- Flom, R., Deák, G. O., Phill, C., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, 27, 181-194.
- Furrow, D., Nelson, K. & Benedict, H. (1979). Mothers' speech to children and syntactic development: Some simple relationships. *Journal of Child Language*, 6, 423-442.
- Giardino, J., Gonzalez, A., Steiner, M., & Fleming, A.S. (2008). Effects of motherhood on physiological and subjective responses to infant cries in teenage mothers: A comparison with non-mothers and adult mothers. *Hormones and Behavior*, 53, 149-158.
- Gleitman, L. R., Newport, E. L. & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, 11, 43-79.
- Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, 100, 8030-8035.
- Goldstein, M.H. & Schwade, J.A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515-523.
- Goldstein, M.H. & Schwade, J.A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.), *The Oxford Handbook of Developmental and Comparative Neuroscience*. Oxford University Press.
- Goldstein, M.H., Syal, S., & Schwade, J.A. (submitted). *Socially guided statistical learning drives the emergence of speech in infants*. Manuscript submitted for publication.
- Goldstein, M.H., Schwade, J.A., & Kirkpatrick, F. (2013, April). *Maternal responsiveness to socially-directed prelinguistic vocalizations predicts later language development*. Paper

presented at the biennial meeting of the Society for Research in Child Development,  
Seattle, WA.

Goldstein, M.H., Waterfall, H.R., Lotem, A., Halpern, J.Y., Schwade, J.A., Onnis, L., &  
Edelman, S. (2010a). General cognitive principles for learning structure in time and place.  
*Trends in Cognitive Sciences, 14*, 249-258.

Goldstein, M. H., Schwade, J. A., Briesch, J., & Syal, S. (2010b). Learning while babbling:  
prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy, 15*,  
362-391.

Goldstein, M.H., & West, M.J. (1999). Consistent responses of human mothers to prelinguistic  
infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology*,  
*113*, 52-57.

Green, J.A., Jones, L.E., & Gustafson, G.E. (1987). Perception of cries by parents and  
nonparents: Relation to cry acoustics. *Developmental Psychology, 23*, 370-382.

Gros-Louis, J., West, M.J., Goldstein, M.H., & King, A.P. (2006). Mothers provide differential  
feedback to infants' prelinguistic sounds. *International Journal of Behavioral*  
*Development, 30*, 112-119.

Gustafson, G. E. & Green, J.A. (1989). On the importance of fundamental frequency and other  
acoustic features in cry perception and infant development. *Child Development, 4*, 772-  
780.

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young*  
*American children*. Brookes: Baltimore, MD.



- Illman, G., Neuhauserova, K., Pokorna, Z., Chaloupkova, H., & Simeckova, M. (2008). Maternal responsiveness of sows towards piglet's screams during the first 24 hours postpartum. *Applied Animal Behaviour Science*, 112, 248-259.
- Kitchen, D.M., Bergman, T.J., Cheney, D.L., Nicolson, J.R., & Seyfarth, R.M. (2010). Comparing responses of four ungulate species to playbacks of baboon alarm calls. *Animal Cognition*, 13, 861-870.
- Landry, S. H., Smith, K. E., Miller-Loncar, C. L., & Swank, P. R. (1997). Predicting cognitive-language and social growth curves from early maternal behaviors in children at varying degrees of biological risk. *Developmental Psychology*, 33, 1040 –1053.
- Ma, E., Lau, J., Grattan, D.R., Lovejoy, D.A., & Wynne-Edwards, K.E. (2005). Male and female prolactin receptor mRNA expression in the brain of a biparental and a uniparental hamster, *Phodopus*, before and after the birth of a litter. *Journal of Neuroendocrinology*, 17, 81-90.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 21, 746-748.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.
- Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of vocal development and disorders. In C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.), *Phonological development* (pp. 509-536). Parkton, MD: New York Press.
- Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, 32, 223–246.

- Oller, D. K., Eilers, R. E., & Basinger, D. (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science*, 4, 49-60.
- Paavola, L., Kunnari, S., Moilanen, I., & Lehtihalmes, M. (2005). The functions of maternal verbal responses to prelinguistic infants as predictors of early communicative and linguistic development. *First Language*, 25, 173–195.
- Papousek, M. (1989). Determinants of responsiveness to infant vocal expression of emotional state. *Infant Behavior and Development*, 12, 507-524.
- Papoušek, H. & Bornstein, M. H. (1992). Didactic parental support for human infants. In H. Papoušek, U. Jürgens, and M. Papoušek (Eds), *Nonverbal Vocal Communications* (pp. 209-229). New York: Cambridge University Press.
- Parke, R. D. (2000). Father involvement: A developmental psychological perspective. *Marriage & Family Review*, 29, 43-58.
- Paul, R. & Jennings, P. (1992) Phonological behavior in toddlers with slow expressive language development. *Journal of Speech & Hearing Research*, 35, 99-107.
- Rescorla, L. & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech & Hearing Research*, 39, 153-165.
- Rickert, M. E. & Yu, C. (2010). *Joint attention through the hands: Investigating the timing of object labeling in dyadic social interaction*. Paper presented at the IEEE 9<sup>th</sup> International Conference on Development and Learning (ICDL), 114-119.
- Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, 24, 221-234.
- Rothbaum, F., & Crockenberg, S. (1995). Maternal control and two-year-olds' compliance and defiance. *International Journal of Behavioral Development*, 18, 193–210.

- Shizawa, Y., Nakamichi, M., Hinobayashi, T., & Minami, T. (2005). Playback experiment to test maternal responses of Japanese macaques (*Macaca fuscata*) to their own infant's call when the infants were four to six months old. *Behavioural Processes*, 68, 41-46.
- Smith, V.A., King, A.P., & West, M.J. (2000). A role of her own: female cowbirds, *Molothrus Ater*, influence the development and outcome of song learning. *Animal Behaviour*, 60, 599-609.
- Smith, N. A. & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, 13, 410-420.
- Stevens, E., Blake, J., Vitale, G. & MacDonald, S. (1998). Mother–infant object involvement at 9 and 15 months: Relation to infant cognition and early vocabulary. *First Language*, 18, 203-222.
- Stoel-Gammon, C. (1992). Prelinguistic vocal development : measurement and predictions. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds), *Phonological development: models, research, implications*, 439–56. Timonium, MD : York Press.
- Tamis-LeMonda, C.S., Bornstein, M.H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72, 748-767.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57, 1454-1463.
- Uvnas-Moberg, K. (1998). Oxytocin may mediate the benefits of positive social interaction and emotions. *Psychoneuroendocrinology*, 23, 819-835.
- van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, 15, 71–74.

- Wood, R.M. & Gustafson, G.E. (2001). Infant crying and adults' anticipated caregiving responses: acoustic and contextual influences. *Child Development*, 72, 1287-1300.
- Yu, C., Smith, L., Fricker, D., & Xu, T. (2013, April). Eyes and hands: The dynamic coordination during free-flowing toy play. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.
- Yu, C. & Ballard, D. H. (2007) A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70, 2149-2165.
- Zeskind, P.S., Klein, L., Marshall, T. R. (1992). Adults' perceptions of experimental modifications of durations and pauses and expiratory sounds in infant crying. *Developmental Psychology*, 28, 1153-1162.
- Zeskind, P.S. & Marshall, T.R. (1988). The relationship between variations in pitch and maternal perceptions of infant crying. *Child Development*, 59, 193-196.

## CHAPTER 5

### CONCLUSIONS

In this thesis, I took a systems approach to communicative development to consider both infants' and caregivers' contributions to the caregiver-infant interaction in developing communicative alignment. Infants typically exist in an environment in which caregivers are highly responsive to infants' vocalizations and infants readily learn from their caregivers' contingent responses to their vocalizations. Infants rapidly change their vocalizations and learn new phonological patterns in response to caregiver input (Goldstein & Schwade, 2008; Goldstein, Syal, & Schwade, under review). While these previous experiments made important contributions to understanding infant vocal development, they left several unanswered questions. When considering the infant side of the dyad, it was unknown what infants were actually learning when they learned new phonological patterns from contingent input to their babbling. Second, the reasons why contingent feedback to babbling facilitates infant learning were understudied. When considering caregivers' contributions to caregiver-infant communication it was unknown which characteristics of infant vocalizations influence caregiver responsiveness and how caregiver responsiveness changes with experience. This dissertation investigated each of these questions to better understand the function of immature sounds in constructing social interactions that facilitate advances in language development.

Chapter 2 built upon the findings of previous studies that demonstrated that infant phonological development takes place through socially guided statistical learning. While previous studies demonstrated that infants use feedback from caregivers to guide their vocal production, it was unclear at what level infants learn phonological regularities. In other words

what exactly is being learned when infants produce new phonological and at what level does learning take place. The results of Chapter 2 suggest that when infants use feedback from caregivers that is contingent on their babbling to learn how to produce well-formed syllables they learn at the level of an acoustic pattern.

In addition to learning phonological patterns from caregiver input to babbling, infant word learning is also facilitated when caregivers' provide object labels contingently upon infants' object-directed vocalizations (Goldstein et al., 2010). Chapter 3 investigated the function of infants' object-directed vocalizations in creating a state of anticipatory readiness. I hypothesized that when infants produce an ODV, the acoustic feedback of hearing their own vocalizations increases stimulation that prepares infants for learning. The results suggest that at least for girls, ODVs can serve to prepare infants for learning as a form of vocal self-stimulation. If infants were actually in a state that is more conducive to learning immediately after vocalizing than at other times it follows that contingent feedback to infant's vocalizations would have powerful effects on infant learning. This result complements previous findings by providing a mechanistic explanation for why infants may be receptive to contingent feedback to their vocalizations.

While caregivers' contingent responses to prelinguistic vocalizations were shown to have both immediate and long-term effects on infants' speech and language development (Goldstein & Schwade, 2010; Tamis-LeMonda, Bornstein, & Baumwell, 2001), much less was known about which characteristics of infant vocalizations drive caregiver responsiveness. Chapter 4 shifted towards a systems approach to development and examined the effects of infant vocalizations on caregiver responsiveness using a playback paradigm. Caregivers reacted to prerecorded audiovisual playbacks of unfamiliar infants' vocalizations and actions. The experiments presented in Chapter 4 were the first to manipulate the acoustic quality and directedness of infant

vocalizations to systematically examine their effects on caregivers' responses. Experiment 1 validated the playback paradigm as a measure of caregiver responsiveness. Experiment 2 demonstrated that caregivers were influenced by the quality and directedness of infant vocalizations when deciding if and how to respond. Caregiver parity also influenced caregivers' perceptions of vocalizations and the types of responses caregivers provided. This investigation of the effects of prelinguistic vocalizations on caregivers' behavior improved our understanding of the function of immature sounds in constructing social interactions that facilitate advances in vocal learning.

### **Specifying the Units of Socially Guided Statistical Learning**

Caregivers consistently provide contingent feedback to infants' vocalizations (Goldstein & West, 1999; Chapter 4). Such consistent and contingent feedback affords infants opportunities to learn the statistical regularities of caregiver input (Goldstein & Schwade, 2008). When exposed to rare phonological patterns, infants rapidly learn and generalize the regularities of the patterns if provided with contingent and variable examples of the phonological pattern (Goldstein, Syal, & Schwade, under review). This process of extracting phonological regularities from caregivers' contingent feedback to infant vocalizations is known as socially guided statistical learning (SGSL; Goldstein & Schwade, 2008; 2010). SGSL has been shown to facilitate infant vocal development. However, much less was known about what constraints may limit infants' vocal production and at what level SGSL takes place. Do infants learn phonological patterns in terms of the auditory patterns present in the phonological input or do they learn in terms of the motor patterns necessary to reproduce the input?

The experiment presented in Chapter 2 tested infants' abilities to learn a novel phonological pattern that did not require a mandibular oscillation to produce. Infants increased

their production of vocalizations that followed the novel phonological pattern demonstrating that infants learn phonology at the level of the acoustic pattern. The results of this experiment provide insight into the limits on SGSL. Several domain-general constraints such as limitations of attention and memory were previously identified for statistical learning of phonological patterns in speech perception (Newport & Aslin, 2004; Toro, Sinett, & Soto-Faraco, 2005). The results of Chapter 2 suggest that infants were not constrained by a dominant mandibular oscillation pattern when producing new phonological patterns. Socially guided statistical learning is therefore a mechanism that allows infants to detect patterns in the social input they receive, extract the statistical regularities of that input as an acoustic pattern, and then generalize knowledge of those patterns by reproducing phonological utterances that match the regularities of the learned acoustic pattern.

### **Anticipatory Readiness as a Mechanism for Infant Learning**

The results of Chapter 3 demonstrated that infants' object-directed vocalizations prepare infants for learning by creating a state of anticipatory readiness. Infant girls who heard their own prerecorded vocalization prior to object labeling learned the word-object association as well as infants who received the label after producing a spontaneous object-directed vocalization (in Goldstein et al., 2010). Girls also learned better than infants who heard the object labeled following a silent look to the object. The results of Chapter 3 help to explain why infants show facilitated learning after contingent parent input. At least for girls, the act of vocalizing appears to induce a state of anticipatory readiness that makes infants more receptive to input after they vocalize. Infants therefore contribute to their own communicative development through vocalizing.



While previous work suggested that word recognition is facilitated when objects are labeled in a full sentence rather than in isolation (Fernald & Hurtado, 2006), the results of Chapter 3 are the first to demonstrate that a single isolated syllable heard prior to labeling also facilitates learning. According to the social-gating hypothesis, social interaction organizes infant attention and increases arousal, thereby facilitating learning (Kuhl, 2007). The results support the concept of social gating by suggesting that the infant's own speech-like sounds create a readiness to learn.

### **Implications of Prelinguistic Communicative Alignment for the Development of Communication and Language**

The results of the experiments presented here also have broader implications for the development of communication. Past research demonstrates that caregivers respond to infants' vocalizations and actions in ways that help organize infant attention (e.g. Baldwin, 1991; Brand, Baldwin, Ashburn, 2002; Deák et al., 2000). Parental behaviors that organize infant attention create socially coordinated, *aligned* interactions in which attentional focus, object movements, and labeling are synchronized in ways that promote learning. Alignment is a process in which communicators' production and comprehension of speech become coupled to establish common ground (Gambi & Pickering, 2011; Pickering & Garrod, 2004). For example, over the course of a dialogue, adult speakers increasingly mirror each other's body language, mannerisms, linguistic conventions, and even syntactic forms (Pickering & Garrod, 2006). One mechanism hypothesized to facilitate communicative alignment is priming (Garrod & Pickering, 2004). Through priming, representations in the listener are activated by the information presented by the speaker. Such priming leads the listener to respond with speech that mirrors the forms used by the speaker.

The development of alignment has been examined only from the caregiver's point of view (e.g. Cohen & Tronick, 1998; Papoušek & Bornstein, 1992). By focusing on the function of infant ODVs, the results of Chapter 3 provide insight into how infants can contribute to the development of aligned social interactions. Infants' ODVs may serve to coordinate the attention of both the infant and the caregiver in triadic interactions with objects. The results of Chapter 3 suggest that ODVs prime infants for learning from caregiver input, but ODVs may also encourage responsiveness in the caregiver. The results of Chapter 4 demonstrate that caregivers respond significantly more to infant vocalizations when they are directed at objects than in an undirected context. Therefore, ODVs may aid in establishing a common ground for communication with the infant by bringing the caregiver's attention to the object at which the vocalization was directed.

Infants' prelinguistic vocalizations also differentially attract caregivers' attention based on the infraphonological qualities of the vocalizations. In the playback experiments presented in Chapter 4, caregivers verbally responded to a higher proportion of more advanced (i.e. fully resonant) vowel sounds. Caregivers also provided different types of information to infants depending on the quality of consonant-vowel syllables. Caregivers were more likely to narrative infants' behaviors after infants produced a marginal syllable while canonical syllables were more likely to elicit imitative responses from caregivers. Thus, prelinguistic vocalizations appear to organize caregiver attention and caregivers structure their responses based on the quality of infants' prelinguistic vocalizations.

A final factor that may influence the development of communicative alignment is caregiving experience. With increasing parity, caregivers' perceptions of the speech-like qualities of prelinguistic vocalizations change. Multiparous women for example, rated canonical

syllables as more speech-like than nulliparous women. Multiparous women also did not rate fully resonant vowels differently from marginal syllables, while nulliparous women rated fully resonant vowels as more speech-like than marginal syllables. This result suggests that multiparous mothers recognized that producing consonants represents an advance in vocal development whereas nulliparous women did not. Parity also influenced the types of responses caregivers provide to vocalizations. Nulliparous women were less likely to imitate infant vocalizations than mothers. Imitation by caregivers supports communicative alignment by promoting social affiliation and prosocial behavior (van Baaren et al. 2004; Chartrand & Bargh, 1999). For example, 18-month-old infants who are imitated are more likely to help others complete an activity than infants whose behavior is not mimicked (Carpenter, Uebel, & Tomasello; 2013). Mothers imitated infant vocalizations more than nulliparous women, particularly the most speech-like vocalizations, perhaps as a way of increasing social affiliation and communicative alignment. Responsiveness to infant vocalizations therefore develops as caregivers gain experience with the various types of vocalizations infants produce and align their interactions with infants' behaviors.

### **Overall Significance**

By studying the infant-caregiver dyad as a functional unit, the results of this dissertation demonstrate that immature vocalizations construct social interactions that facilitate vocal development and language learning. Infants actively contribute to their own language development by learning acoustic patterns through socially guided statistical learning and stimulating a readiness to learn via object-directed vocalizations. At the same time, the qualities and directedness of infants' vocalizations organize caregivers' attention and influence the ways in which they respond.

## References

- Baldwin, D. A. (1991). Infants' contribution to achievement of joint reference. *Child Development*, 62, 874-890.
- Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, 5, 72-83.
- Carpenter, M., Uebel, J., & Tomasello, M. (2013). Being Mimicked Increases Prosocial Behavior in 18-Month-Old Infants. *Child Development*.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893-910.
- Cohen, J. F. & Tronick, E. Z. (1998). Mother-infant face to face interaction: Influence is bi-directional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology*, 24, 386-392.
- Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36, 511-523.
- Fernald, A. & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9, F33-F40.
- Gambi, C. & Pickering, M. J. (2011). A cognitive architecture for the coordination of utterances. *Frontiers in Psychology*, 2, 1-14.
- Garrod, S. & Pickering, M. J. (2004). Why is conversation so easy? *Trend in Cognitive Sciences*, 8, 8-11.
- Goldstein, M. H., & Schwade, J. A. (2010). From birds to words: Perception of structure in social interactions guides vocal development and language learning. In M. S. Blumberg, J. H.

- Freeman, & S. R. Robinson (Eds.), *The Oxford handbook of developmental behavioral neuroscience* (pp. 708–729). Oxford: Oxford University Press.
- Goldstein, M. H., & Schwade, J. A. (2008). Social interaction to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*, 515–523.
- Goldstein, M. H., & West, M. J. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology, 113*, 52–58.
- Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants association their own noncry vocalizations with responses from caregivers. *Child Development, 80*, 636-644.
- Goldstein, M.H., Syal, S., & Schwade, J.A. (submitted). *Socially guided statistical learning drives the emergence of speech in infants*. Manuscript submitted for publication.
- Goldstein, M. H., Schwade, J. A., Briesch, J., & Syal, S. (2010). Learning while babbling: prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy, 15*, 362-391.
- Kuhl, P.K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science 10*, 110-120.
- Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.
- MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*, 499-546.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Papoušek, H. & Bornstein, M. H. (1992). Didactic parental support for human infants. In

- Papoušek, (Eds), *Nonverbal Vocal Communications* (pp.). New York; Cambridge University Press.
- Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-189.
- Pickering, M. J. & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language & Computation*, 4, 203-228.
- Tamis-LeMonda, C.S., Bornstein, M.H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, 72(3), 748-767.
- Toro, J.M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97, B25-B34.
- van Baaren, R. B., Holland, R. W., Kawakami, K., & van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, 15, 71-74.